Lund University

Master's Thesis

On Modelling Extreme Foreign Exchange Volatility Using Copulas

Authors:
Henrik Bosaeus
Anna Sandström

Supervisor:
Nader Tajvidi

June 4, 2015

Lund University

LUND UNIVERSITY

Abstract

On Modelling Extreme Foreign Exchange Volatility Using Copulas

by Henrik Bosaeus and Anna Sandström

The price volatility is an important property to monitor in financial trading. A volatile period implies threats of large losses, but at the same time opportunities of higher gains. This makes accurate volatility prediction models an important part of an algorithmic trading system. This thesis work investigates the extreme value dependence between the Foreign Exchange rate volatility and the rate of change of the offer- and bid volumes present at the market. Each of the currency pairs EURUSD, EURSEK and EURNOK will be analyzed for a one hour period on three different days, where trading volumes and prices are given at several levels at each point in time. The first part of the thesis aims to transform the data into one-dimensional data series, describing the rate of fluctuation of price and volume. These are then subject to time series model fitting in order to remove all forms of autocorrelations in the data. The extreme values of the residual series are then extracted using a block maxima approach, and modelled using the Generalized Extreme Value (GEV) distribution. The quite novel approach of extreme value copulas is then applied as a method of modelling the joint dependence structure between the volume- and price extremes. The results indicate dependence being present, in most cases appropriately described by extreme value copula models. Further research on the topic is suggested.

Key words: Foreign Exchange, Volatility, ARMA-GARCH, Block Maxima, Extreme Values, Copula, Maximum Likelihood.

Acknowledgements

We would like to thank our supervisor Nader Tajvidi at the Faculty of Engineering, Lund University, for his support and valuable statistics expertise. We are also very grateful to Pär Hellstrom at SEB for introducing us to the subject and for his helpful opinions throughout the project.

Contents

Chapter 1

Introduction

1.1 Aim and scope of thesis

Statistics is one of several theoretical areas which might be usable for developing algorithmic models for high frequency Foreign Exchange (FX) rate trading. Predictive models for FX rate volatility, being a measure of its fluctuation, is of special importance as a high volatility provides both an opportunity but also a threat to the trader. External inputs to the prediction model might be any external information for which there is a causal relationship to the volatility. The purpose of this thesis is to investigate whether the volumes that buyers and sellers are willing to trade at the market might provide valuable information for predicting the price volatility. More specifically, dependence between extremes in the degree of fluctuation of the volume and price will be investigated, mainly through statistical methods of time series and extreme value copulas. If the results show significant dependence between the variables, further research might be recommended in order to expand existing trading models to include the new variable.

Three currency pairs will be investigated, namely EURSEK, EURUSD and EURNOK. Using this notation, buying for example EURSEK means that SEK is used to buy EUR. As EUR is a stronger currency than SEK, the rate EURSEK is greater than one. Each pair will be investigated for a one hour interval on three different days and, due to confidentiality reason and, the exact time and date will not be mentioned. Data regarding the prices and volumes at each point in time will be subject to a series of transformations, resulting in univariate time series describing the price- and volume fluctuations at each point in time. Two series will be produced from each original series, corresponding to different sampling speeds for extracting data from the original series. After removing auto correlations through time series modelling of the price- and volume series, the extreme values of their residuals will be extracted using a block maxima approach. These extremes will then be modelled, first separately using univariate extreme value models, then together using extreme value copulas. Finally, conditional probability plots will be produced for inference regarding any possible dependence between the variables.

1.2 The FX market

The FX market is the world's largest financial market, with an average daily turnover of more than 2 trillion U.S. Dollars [Weithers, 2006]. With market participants ranging from banks and large corporations to individuals, the market is available to anyone wanting to speculate in currency movements.

The way in which buyers and sellers exchange currencies has coevolved with the technological progress. In earlier days even smaller transactions had to be carried out manually, a process being error-prone, expensive and with costs being passed on to the client [Aldridge, 2010]. When electronic dealing systems were introduced in the 1980s, these were unsurprisingly adopted and heralded as revolutionary. Since then, technical development has increased transaction speed and the way in which the market is supplied with information. By enabling computerized systems to make predictions and conduct transactions within a short time frame, the concept of high frequency trading has grown prominent, today accounting for the major trading volume on the FX market [Aldridge, 2010].

While traditional FX traders may hold their trading positions for weeks, the corresponding time frame for high frequency traders may be in the order of microseconds. Being fully automatic, high frequency trading is characterized by being active around the clock, making frequent transactions but with a relatively low average gain per trade [Aldridge, 2010].

Chapter 2

Theory

2.1 Volatility estimation

One method of estimating the volatility is to use an *Exponentially Weighted Moving Average (EWMA)*. Using this method, each of the previous observations is assigned a weight getting exponentially smaller the further the distance in time from the volatility value being evaluated. If the time series is zero mean, the equation reduces to

$$\sigma_t^2 = (1-\lambda)\sum_{i=1}^{n} \lambda^{i-1} x_{t-i}^2 \tag{2.1}$$

where λ is a "forgetting constant" between 0 and 1, σ_t^2 the squared volatility and x_{t-i} the time series value at time $t-i$, in this case being the log-return of either price or volume. The smaller the value of λ, the larger is the relative weight given to recent samples [Dowd, 2005].

2.2 Spacing tick data

One way to overcome the irregularities of tick data is to sample it at regular periods in time. In order to achieve a sample estimate at equally spaced time points linear time weighted interpolation can be used. This technique estimates unobserved samples as lying on a straight line that connects two neighbouring observed samples, using the formula

$$\hat{x_t} = x_{t,last} + (x_{t,last} - x_{t,next})\frac{t - t_{last}}{t_{next} - t_{last}}$$

where $\hat{x_t}$ is the resulting sample, $x_{t,last}$ and $x_{t,next}$ the previous and following tick data sample, t the desired sampling time and t_{last} and t_{next} the time stamps of the previous and following tick data sample [Aldridge, 2010].

2.3 ARIMA-(E)GARCH processes

The *Autoregressive Moving Average* (ARMA) process of order (m,n) is defined as

$$y_t = \sum_{i=1}^{m} a_i y_{t-i} + \varepsilon_t + \sum_{j=1}^{n} c_j \varepsilon_{t-j} \tag{2.2}$$

where y_t is the time series value and ε_t a zero-mean white noise process. The ARMA model thus considers the correlation between samples at different time points in the data. If the time series has a stochastic trend, or if each value is strongly dependent on the previous, initial differentiation of the data is suitable, according to

$$\nabla y_t = y_t - y_{t-1}$$

Fitting an ARMA(m,n) model to differentiated data gives an *Autoregressive Integrated Moving Average* (ARIMA) process of order (m,d,n) with d being the number of differentiations of the data [Jakobsson, 2013].

Further, time series can suffer from heteroscedasticity meaning that the variance of the time series changes over time. By modelling the innovations, ε_t, as a product of a time variant non-stochastic part σ_t, and a stochastic part z_t, being a white noise process with mean 0 and variance 1,

$$\varepsilon_t = \sigma_t z_t \tag{2.3}$$

the heteroscedacity can be accounted for.

The *Generalized Auto Regressive Conditional Heteroskedasticity* (GARCH(p,q)) models model the time variant volatility by using the above representation of the innovations and estimating σ_t as

$$\sigma_t^2 = \omega + \sum_{k=1}^{p} \alpha_k \varepsilon_{t-k}^2 + \sum_{l=1}^{q} \beta_l \sigma_{t-l}^2 \tag{2.4}$$

with $\omega > 0$, $\alpha_k > 0$ for $k = 1...p$, $\beta_l > 0$ for $l = 1...q$ and $\sum_k^p \alpha_k + \sum_l^q \beta_l < 1$ to ensure stability [Wurtz et al.].

As a negative innovation might, in some cases, have a different impact on the variance than a positive innovation of the same magnitude, the GARCH model has the limitation of being symmetric. The *Exponential GARCH (EGARCH) process* addresses this problem, at the same time imposing less restrictions on the parameters. Instead forming σ_t according to

$$\ln(\sigma_t^2) = \omega + \sum_{k=1}^{p} (\alpha_k z_{t-k} + \gamma_k(|z_{t-k}| - E|t_{t-k}|)) + \sum_{l=1}^{q} \beta_l \ln(\sigma_{t-l}^2)$$

the α parameter now captures the sign effect and γ the size effect [Ghalanos, 2014].

2.4 Residual distributions

The errors of a time series model, being the difference between expected and observed values, are also called residuals. In the process of optimizing parameter estimates for a given time series model, an assumption is being made regarding the type of distribution of the standardized residuals, being referred to as z_t in equation (2.3). In choosing an appropriate standardized distribution to fit to the residuals, the skewness and kurtosis are important properties to consider.

Skewness
Being a measure of asymmetry of the distribution, the skewness of variable X is often measured according to

$$skewness = \frac{\mathrm{E}\left[(X-\mu)^3\right]}{\sigma^3}$$

with E denoting the expected value, μ the mean and σ the standard deviation of the sample. If the skewness is zero, the distribution is symmetric. Similarly, if the skewness is positive the distribution has a long right tail and if the skewness is negative, it has a long left tail.

Kurtosis
The peakiness of the distribution is referred to as its kurtosis, β, commonly measured according to

$$\beta = \frac{\mathrm{E}\left[(X-\mu)^4\right]}{\sigma^4}$$

The normal distribution has a kurtosis of 3, hence if $\beta > 3$ the distribution is more peaked than the normal distribution and if $\beta < 3$ the distribution is more flat than the normal distribution [Dowd, 2005].

2.4.1 Generalized error distribution (GED)

For many decades, it was assumed that the error random variable follows a normal distribution with mean zero. As this assumption was found to be incorrect in some cases, Subbotin (1923) introduced the symmetric class of generalized error distributions (GED), with similar structural properties but with varying kurtosis. This distribution is also called generalized normal distribution or exponential power distribution [Vasudevay and Kumari, 2013].

The GED is a three parameter distribution with conditional density given by

$$f(x) = \frac{\kappa e^{-0.5\left|\frac{x-\alpha}{\beta}\right|^{\kappa}}}{2^{1+\kappa^{-1}}\beta\Gamma(\kappa^{-1})} \tag{2.5}$$

with α, β and κ representing the location, scale and shape parameters respectively. Since the distribution is symmetric, the location parameter α is equal to the mean μ. The GED variance is given by

$$\mathrm{V}(x) = \beta^2 2^{2/\kappa}\frac{\Gamma(3\kappa^{-1})}{\Gamma(\kappa^{-1})}$$

Standardization of this distribution is a location scale transformation, using that two distribution functions G_1 and G_2 are of the same distribution type if $G_1(x) = G_2(\sigma x + \mu)$. The transformation is performed by scaling and solving the case $\mathrm{V} = 1$ for β. This gives

$$\beta = \sqrt{2^{-2/\kappa}\frac{\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})}}$$

By substitution of β in equation 2.5, the density of the standardized variable $z = \frac{x-\mu}{\sigma}$ can finally be yielded as

$$f\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}f(z) = \frac{1}{\sigma}\frac{\kappa e^{-0.5\left|\sqrt{2^{-2/\kappa}\frac{\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})}}z\right|^{\kappa}}}{\sqrt{2^{-2/\kappa}\frac{\Gamma(\kappa^{-1})}{\Gamma(3\kappa^{-1})}}2^{1+\kappa^{-1}}\Gamma(\kappa^{-1})}$$

where μ is the mean of x and σ^2 the variance of the variable. In this work the GED distribution is assumed for the ARIMA-GARCH model, in which case σ is the conditional heteroscedasticity of equation (2.3) [Ghalanos, 2014].

2.4.2 Skewed generalized error distribution (SGED)

In order to account for asymmetry of the residuals, Fernandez and Steel (1998) proposed a way to skew a symmetric distribution. Given a skew parameter ξ, the density of a random variable z can be represented as

$$f(z|\xi) = \frac{2}{\xi + \xi^{-1}} \left[f(\xi z) H(-z) + f(\xi^{-1} z) H(z) \right] \tag{2.6}$$

with $\xi \in \mathbb{R}^+$ and $H(.)$ being the Heaviside function. The absolute moments, used to derive central moments such as the mean and variance, is given by

$$M_r = 2 \int_0^\infty z^r f(z) dz \tag{2.7}$$

and by using the two functions of equation (2.6) and (2.7), the standardized skewed generalized error distribution (SGED) can be created, having zero mean and unit variance [Ghalanos, 2014].

2.5 Generalized Extreme Value distribution

Consider the sequence $X_1, ..., X_n$ of independent random variables, having a common distribution function F. Forming the stochastic maximum of this block of n values,

$$M_n = \max(X_1, ..., X_n)$$

the distribution function of M_n can be created as

$$\begin{aligned} \mathrm{P}(M_n \leq z) &= \mathrm{P}(X_1 \leq z, ..., X_n \leq z) \\ &= \mathrm{P}(X_1 \leq z) \times ... \times \mathrm{P}(X_n \leq z) \\ &= F(z)^n \end{aligned}$$

While it seems simple to estimate this distribution by estimating $F(z)$, then taking it to the power of n, small discrepancies in the estimate of F may cause large discrepancies for F^n. This problem might be avoided by finding an estimate for F^n directly. By considering the behaviour of F^n as $n \to \infty$, modelling seems problematic as M_n degenerates to a point at the upper endpoint of F. But, by renormalizing M_n by

$$M_n^* = \frac{M_n - b_n}{a_n}$$

for appropriate sequences of constants $\{a_n > 0\}$ and $\{b_n\}$, appropriate choices of constants might stabilize the location and scale of M_n^* for increasing n. This leads to the follow theorem.

Theorem 2.5.1. If there exist sequences of constants $a_n > 0$ and b_n such that

$$\mathrm{P}((M_n - b_n)/a_n \leq z) \to G(z) \quad \textit{as } n \to \infty$$

for a non-degenerate distribution function G, then G is a member of the GEV family

$$G(z) = \exp\left\{-\left[1 + \xi(\frac{z-\mu}{\sigma})\right]^{-1/\xi}\right\},$$

defined on $z : 1 + \xi(z - \mu)/\sigma > 0$, where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$. □

Now, assuming that

$$\mathrm{P}((M_n - b_n)/a_n \leq z) \approx G(z)$$

as $n \to \infty$, it follows that

$$\begin{aligned} \mathrm{P}(M_n \leq z) &\approx G((z - b_n)/a_n) \\ &= G^*(z) \end{aligned}$$

where G^* is also a member of the GEV family. Therefore there is no need to first approximate the parameters of the renormalized maxima, but the parameters of the actual block maxima M_n can be estimated directly [Coles, 2001].

2.6 Copulas

The word copula is Latin and means "connect" or "join", which is suitable since a copula is a function that connects univariate distribution functions into multivariate distributions. This can for instance be useful when modelling dependence between random variables. Copulas are a quite new phenomena but has gained popularity within fields of applied mathematics, like finance, insurance and reliability theory [Durante et al., 2009].

2.6.1 Sklar's theorem

The essentials of copulas are summarized in Sklar's theorem below.

Theorem 2.6.1 (Sklar's theorem)**.** Let H be a bivariate distribution function with margins F and G. Let $RanF$ and $RanG$ denote the range of F and G. Then there exists a copula C such that for all x, y in $\mathbb{R}$,

$$H(x, y) = C(F(x), G(y))$$

then C is uniquely determined on $RanF \times RanG$ if F and G are continuous. Conversely, if C is a copula and F and G are distribution functions, then the function H is a joint distribution function with margins F and G [Nelson, 2006].

2.6.2 Necessary properties

Further, let $\mathbb{I} = [0, 1]$, then a two-dimensional copula C is a function $\mathbb{I} \to \mathbb{I}^2$ if it satisfies the following properties: For every u,v in $\mathbb{I}$,

$$C(u,0) = 0 = C(0,v)$$
$$C(u,1) = u \quad \text{and} \quad C(1,v) = v$$

For every u_1, u_2, v_1, v_2 in $\mathbb{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

It can also be shown that every copula C is bounded by,

$$W(u,v) = \max(u + v - 1, 0) \leq C(u,v) \leq \min(u,v) = M(u,v)$$

[Nelson, 2006].

2.6.3 Extreme value copulas

In order to model dependence of multivariate extremes, a special family of copulas may be used, called extreme value copulas. The formal definition of an extreme value copula C_* is that if there exist a copula C such that

$$C_*(u,v) = \lim_{n\to\infty} C^n(u^{1/n}, v^{1/n})$$

for u, v in $\mathbb{I}$, then C_* is an extreme value copula [Nelson, 2006]. In this thesis work, five different extreme value copulas will be used to model bivariate dependence. The parameter θ is in all definitions the copula parameter, showing the level of dependence between the variables.

It has also been shown that an extreme value copula can be expressed in the following form

$$C(u,v) = \exp\left[\log(uv)A\left(\frac{\log(u)}{\log(uv)}\right)\right]$$

where $A : [0,1] \to [1/2,1]$ is Pickand's dependence function. $A(t)$ is a convex function such that $\max(t, 1-t) \leq A(t) \leq 1 \,\forall\, t \in [0,1]$. The two limits of A has special meanings, where the upper bound $A(t) = 1$ stands for independence corresponding to $C(u,v) = uv$, and the lower bound $A(t) = \max(t, 1-t)$ corresponds to perfect dependence [Durante et al., 2009].

Tawn copula

The Tawn copula is defined as

$$C(u,v) = uv\exp\left(-\theta\frac{\log(u)\log(v)}{\log(uv)}\right)$$

with $0 \leq \theta \leq 1$ describing the dependence between the variables. As $\theta \to 1$, the dependence increases, however perfect dependence can not be accomplished.

Husler-Reiss copula

The Husler-Reiss copula is defined as

$$C(u,v) = \exp\left(-\hat{u}\ \Phi\left[\frac{1}{\theta} + \frac{\theta}{2}\log\left(\frac{\hat{u}}{\hat{v}}\right)\right] - \hat{v}\ \Phi\left[\frac{1}{\theta} + \frac{\theta}{2}\log\left(\frac{\hat{v}}{\hat{u}}\right)\right]\right)$$

where $0 \leq \theta < 1$, $\hat{u} = -\log(u), \hat{v} = -\log(v)$ and Φ is the standard normal distribution.

The dependence function is given by

$$A(t) = t\Phi\left[\frac{1}{\theta} + \frac{1}{2}\theta\log\left(\frac{t}{1-t}\right)\right] + (1-t)\Phi\left[\frac{1}{\theta} + \frac{1}{2}\theta\log\left(\frac{t}{1-t}\right)\right]$$

where the dependence increases as $\theta \to \infty$ [Demarta and McNeil, 2005].

Galambos copula

The dependence function for the Galambos copula is

$$A(t) = 1 - (t^{-\theta} + (1-t)^{-\theta})^{-1/\theta}.$$

This yields the bivariate copula

$$C(u,v) = uv \exp\left[\left((-\log(u))^{-\theta} + (-\log(v)^{-\theta})\right)^{-1/\theta}\right]$$

where $0 \leq \theta < \infty$. θ equal to 0 indicates independence and the dependence increases as $\theta \to \infty$ [Mathieu and Mohammed, 2012].

Gumbel copula

The dependence function $A(t)$ for the Gumbel copula is defined as

$$A(t) = (t^{\theta} + (1-t)^{\theta})^{1/\theta}$$

This yields the bivariate copula

$$C(u,v) = \exp\left[-\left((-\log(u))^{\theta} + (-\log(v))^{\theta}\right)^{1/\theta}\right]$$

[Nelson, 2006]

Here, $\theta \leq 1$ expresses the degree of dependence, where $\theta = 1$ corresponds to independence and as $\theta \to \infty$, the degree of dependence reaches perfect dependence [Ghorbel and Trabelsi, 2009].

Extremal t copula

The dependence function for the extremal t copula is

$$A(t) = t\ T_{\nu+1}\left(\frac{\left(\frac{t}{1-t}\right)^{1/\nu} - \rho}{\sqrt{1-\rho^2}}\sqrt{\nu+1}\right) + (1-t)T_{\nu+1}\left(\frac{\left(\frac{1-t}{t}\right)^{1/\nu} - \rho}{\sqrt{1-\rho^2}}\sqrt{\nu+1}\right)$$

[Mathieu and Mohammed, 2012]. This yields the copula as

$$\begin{aligned} C(u,v) = \exp\Bigg[& T_{\nu+1}\left\{-\frac{\rho}{b} + \frac{1}{b}\left(\frac{\log(v)}{\log(u)}\right)^{1/\nu}\right\}\log(u) + \\ & T_{\nu+1}\left\{-\frac{\rho}{b} + \frac{1}{b}\left(\frac{\log(u)}{\log(v)}\right)^{1/\nu}\right\}\log(v)\Bigg] \end{aligned}$$

where T_ν is the cumulative distribution function of a Student random variable with ν degrees of freedom, $b^2 = \frac{1-\rho^2}{1+\nu}$ and $\rho \in (-1, 1)$ is the off-diagonal element of the dispersion matrix [Demarta and McNeil, 2005].

2.7 Random variate generation

Random number generation of (u, v) from a bivariate copula $C(u, v)$ is performed in a few easy steps:

1. Create the conditional probability with respect to one of the variables

$$\begin{aligned} \mathrm{P}(V \le v \mid U = u) &= \lim_{\delta \to 0} \frac{\mathrm{P}(u - \delta \le U \le u, V \le v)}{\mathrm{P}(u - \delta \le U \le u)} \\ &= \lim_{\delta \to 0} \frac{C(u,v) - C(u-\delta, v)}{\delta} \\ &= \frac{\partial C(u,v)}{\partial u} \\ &= c_u(v) \end{aligned}$$

2. Generate independent u and t from $U(0,1)$

3. Create v as

$$v = c_u^{-1}(t)$$

The pair (u, v) can now be regarded as an observation from $C(u, v)$ [Nelson, 2006].

2.8 Estimation of copulas

2.8.1 Parameter estimation

One step approach (FML)

Suppose that n independent realizations $\{(X_{i1}, X_{i2})^T : i = 1, ..., n\}$ have been observed from a bivariate distribution. Let the marginal distribution functions be denoted by F_1 and F_2, the corresponding density functions by f_1 and f_2, the vector of all marginal parameters by $\boldsymbol{\beta}$ and the copula parameter by α. The total set of parameters to be estimated is then $\theta = (\boldsymbol{\beta}, \alpha)$.

Now, forming the log likelihood function as

$$l(\theta) = \sum_{i=1}^{n} \log \ c\{F_1(X_{i1}; \boldsymbol{\beta}), F_2(X_{i2}; \boldsymbol{\beta}); \alpha\} + \sum_{i=1}^{n} \sum_{j=1}^{2} \log \ f_i(X_{ij}; \boldsymbol{\beta})$$

the estimate of all parameters simultaneously can then be found as the argument θ maximizing the log likelihood

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmax}} \ l(\theta)$$

with Θ being the parameter space [Yan, 2007]. As all parameters are estimated simultaneously using maximum likelihood this method is referred to as a *Full Maximum Likelihood (FML)* method.

Two step approach (IFM)

A two-stage estimation method was proposed bu Joe and Xu (1996) as an alternative to this method, called inference functions for margins (IFM). Letting $\boldsymbol{\beta_j}$ denote the parameters of F_j, each set of marginal parameters are first estimated using log likelihood maximization according to

$$\hat{\boldsymbol{\beta}}_{jIFM} = \underset{\beta_j}{\operatorname{argmax}} \sum_{i=1}^{n} \log \; f_i(X_{ij}; \boldsymbol{\beta}_j)$$

Given these parameters, the copula parameter θ is estimated as

$$\hat{\theta}_{IFM} = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log \left(c \left(F_1(X_{i1}; \hat{\boldsymbol{\beta}}_{IFM}), ..., F_p(X_{ip}; \hat{\boldsymbol{\beta}}_{IFM}); \theta \right) \right)$$

Since the IFM method optimizes fewer parameters in each step, it provides a computationally less demanding method, which also allows comparison of results after each step [Yan, 2007].

2.8.2 Confidence interval

In order to assess the certainty of a parameter θ_i of the parameter vector θ, confidence intervals can be created at a confidence level of interest.

Profile likelihood method

An accurate method of creating the confidence interval is based on profile likelihood. Denoting the log likelihood of θ as $\ell(\theta_i, \theta_{-i})$, where θ_{-i} denotes all components of θ excluding θ_i, the profile log-likelihood of θ_i is defined as

$$\ell_p(\theta_i) = \max_{\theta_{-i}} \ell(\theta_i, \theta_{-i})$$

The idea is to treat θ_i as fixed and maximize the log likelihood with respect to all other parameters. This is repeated for a scope of values of θ_i. Further, letting $\hat{\theta}$ denote the maximum likelihood estimate of the parameter vector $\theta = (\theta_i, \theta_{-i})$ and using that

$$D_p(\theta_i) = 2(\ell(\hat{\theta}) - \ell_p(\theta_i)) \sim \chi_1^2,$$

an $(1-\alpha)$ interval for θ_i can be constructed by $C_\alpha = \{\theta_i : D_p(\theta_i) \leq c_\alpha\}$ where c_α is the $1-\alpha$ quantile of the χ_1^2 distribution [Coles, 2001].

2.9 Dependence measures

2.9.1 Linear dependence

Pearson's correlation coefficient, ρ_P is a commonly used measure of dependence between two variables X and Y. It is defined as

$$\rho_P(X,Y) = \frac{\mathrm{C}(X,Y)}{\sqrt{\mathrm{V}(X)\,\mathrm{V}(Y)}}$$

where V and C denote the variance and the covariance. Being a linear measure of dependence, $\rho_P(X,Y)$ proves simple and useful for linearly related coefficients, but is insufficient for measuring non-linear relationships. In addition, the measure does not allow any transformation of the variables.

2.9.2 Non-linear dependence

In many cases the relationship between variables is not linear, in which case ρ_P provides a spurious measure. For this purpose we introduce *Kendall's* τ and *Spearman's* ρ, which are based on the concept of *concordance* and *discordance*.

Concordance and discordance

A pair of observations (x_i, y_i) and (x_j, y_j) are said to be concordant if $(x_i - x_j)(y_i - y_j) > 0$ and discordant if $(x_i - x_j)(y_i - y_j) < 0$. A line through the two points will thus have a positive slope if concordant, and negative if discordant.

Kendall's $\boldsymbol{\tau}$

The population version of this dependence measure (τ_C) is the difference between the probability of concordance and the probability of discordance,

$$\tau_{X,Y} = \mathrm{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathrm{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

where (X, Y) are continuous random variables. This difference can be expressed analytically using copulas, as

$$\tau_C = 4 \int\int_{\mathbb{I}^2} C(u,v)dC(u,v) - 1$$

where $C(u, v)$ couples the distribution functions of X and Y.

The sample version of Kendall's τ can easily be estimated from an observation $(x_1, y_1), ..., (x_n, y_n)$ of the variables. Letting c and d denote the number of concordant and discordant pairs respectively, the sample estimate is given by

$$\tau_K = \frac{c-d}{c+d}$$

[Nelson, 2006]

Spearman's ρ
This measure of dependence provides an alternative to Kendall's τ, with the population version given by

$$\rho_{X,Y} = 3(\mathrm{P}[(X_1 - X_2)(Y_1 - Y_3) > 0] - \mathrm{P}[(X_1 - X_2)(Y_1 - Y_3) < 0])$$

Expressed using copulas,

$$\rho_C = 12 \int\int_{\mathbb{I}^2} uv \; dC(u,v) - 3$$

2.10 Model validation

2.10.1 The auto-correlation function (ACF)

The *Auto Correlation Function* (ACF) is a valuable model validation tool when applied on the residuals of a fitted time series model.

First, the auto covariance will be introduced, being a measure of correlation within a time series. This will contribute to understanding of the ACF. The auto-covariance at lag k is defined as $r_y(k) \equiv \mathrm{C}[y_t, y_{t-k}]$ and is commonly estimated using either an *unbiased* estimation method yielding $\hat{r}_y^u(k)$ or an asymptotically unbiased estimation method yielding $\hat{r}_y^b(k)$, according to

$$\hat{r}_y^u(k) = \frac{1}{N-k} \sum_{t=k+1}^{N} (y_t - \hat{m}_y)(y_{t-k} - \hat{m}_y)^T$$
$$\hat{r}_y^b(k) = \frac{1}{N} \sum_{t=k+1}^{N} (y_t - \hat{m}_y)(y_{t-k} - \hat{m}_y)^T$$

where N is the sample size, k is the lag and $\hat{m}_y$ the mean of the sample.

Being very similar, $\hat{r}_y^u(k)$ has the benefit of providing a less biased estimate than $\hat{r}_y^b(k)$. On the other hand, $\hat{r}_y^b(k)$ is always formed by division by N, which in the case of large k reduces the estimated values and hence the variance of the estimates. For large sample sizes, it should be noted that the two methods provide very similar estimates.

The auto-correlation function of y_t is now defined as

$$\rho_y(k) = \frac{r_y(k)}{r_y(0)}, \qquad \text{k} = 0,1,2,...$$

and provides a value easier to interpret. As the magnitude of $r_y(k)$ is greatest for k = 0, the auto-correlation is bounded such that $|\rho_y(k)| \leq 1$.

When evaluating the ACF of the residuals of a time series, these are aimed to be independent and identical distributed variables of a white noise process. Assumed to be zero-mean with variance σ_e^2, the expected value and variance of the auto-correlation estimates are given by:

$$\mathrm{E}(\hat{\rho}_e(k)) = 0$$
$$\mathrm{V}(\hat{\rho}_e(k)) = \frac{1}{N}$$

for k $\neq$ 0. $\hat{\rho}_e(k)$ is asymptotically normal distributed and the confidence interval can thus be created as

$$\hat{\rho}_e(k) \approx 0 \pm \frac{\lambda_\alpha}{\sqrt{N}}, \qquad \mathrm{k} \neq 0$$

By visual assessment, a sequence might be considered to be white noise if a proportion less than α exceeds the confidence interval [Jakobsson, 2013].

2.10.2 Weighted Ljung-Box test

The weighted Ljung-Box test examines the autocorrelations of the residuals under the null hypothesis that $\rho_1 = \rho_2 = ... = \rho_m = 0$. The test statistic is

$$\tilde{Q}_w^* = n(n+2)\sum_{k=1}^{n} \frac{(m-k+1)}{m} \frac{\tilde{r}_k^2}{n-k}$$

where n is the number of observations and m the number of examined lags. The test is performed for both standardized and squared standardized residuals [Gallagher and Fisher, 2013].

2.10.3 Lagrange multiplier test of Engle

The Lagrange multiplier test of Engle tests for ARCH-effects in the data. This can be used both on the residuals after ARIMA-modelling to examine if an ARCH-model must be added to the model, but it can also be used on the residuals from a fitted ARCH-model to see if the ARCH-effects are taken care of properly. The test tries the alternative hypothesis

$$H_1 : \varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \ldots + \alpha_{t-m} \varepsilon_{t-m}^2 + e_t, \qquad t = m+1, \ldots N$$

against the null hypothesis

$$H_0 : \alpha_1 = \alpha_2 = \ldots \alpha_m = 0$$

where e_t is an error term, m a pre specified integer and N is the sample size. To derive the test statistic F, let

$$SSR_0 = \sum_{t=m+1}^{N} \left(a_t^2 - \frac{1}{N} \sum_{i=1}^{T} a_i^2 \right)^2$$

and

$$SSR_1 = \sum_{t=m+1}^{N} \hat{e}_t^2$$

where $\hat{e_t}$ is the residual of the prior linear regression. Then,

$$F = \frac{(SSR_1 - SSR_0)/m}{SSR_1/(N - 2m - 1)}$$

which asymptotically is distributed as χ^2_m under the null hypothesis. If $F > \chi^2_m(\alpha)$ for some confidence level α, the null hypothesis is rejected [**?**].

2.10.4 Sign Bias Test

The sign bias test of Engle and Ng tests the presence of leverage effects in the standardized residuals. In other words if the model should be improved to account for the fact that not only the magnitude but also the sign of the innovations affect the variance estimate. This test captures possible mis-specifications of the GARCH model, by regressing the squared standardized residuals on lagged negative and positive shocks as follows:

$$\hat{z}_t^2 = c_0 + c_1 \mathbb{1}_{\hat{\varepsilon}_{t-1}<0} + c_2 \mathbb{1}_{\hat{\varepsilon}_{t-1}<0} \hat{\varepsilon}_{t-1} + c_3 \mathbb{1}_{\hat{\varepsilon}_{t-1}\leq 0} \hat{\varepsilon}_{t-1} + u_{t-1}$$

where $\mathbb{1}$ denotes the indicator function and $\hat{\varepsilon}_t$ is the residuals from the estimated GARCH-process. Here the null hypothesis is $H_0 : c_i = 0$ for $i = 1, 2, 3$ and jointly that $c_1 = c_2 = c_3 = 0$. If a GARCH model fails the sign bias test this implies that positive and negative shocks have different reactions on the conditional variance. If this is the case, an EGARCH model might be more appropriate [Ghalanos, 2014].

2.10.5 Information criteria

The information criteria of a model is commonly used in model order selection. The value is based on information theory, and penalizes a high model order and at the same rewarding a high likelihood of the model generating the observation. These statistics could be especially useful when making the final judgement about which model to choose. If two or more models perform similarly the one with the lowest information criteria is chosen.

There exists many versions of the information criteria, where the *rugarch* package in R provides the following four: *Akaike (AIC)*, *Bayesian (BIC)*, *Hannan-Quinn (HQIC)* and *Shibata (SIC)*, presented below. Here, N is the number of observations, LL the log likelihood and m the number of unknown parameters estimated [Ghalanos, 2014].

$$\begin{aligned}
AIC &= \frac{-2LL}{N} + \frac{2m}{N} \\
BIC &= \frac{-2LL}{N} + \frac{m\ln(N)}{N} \\
HQIC &= \frac{-2LL}{N} + \frac{2m\ln(\ln(N))}{N} \\
SIC &= \frac{-2LL}{N} + \ln\left(\frac{N+2m}{N}\right)
\end{aligned}$$

2.10.6 Fit diagnostics

When fitting a distribution to a data set, inference should be made regarding the adequacy of the fit. For this purpose, several visual tools can be used.

Letting $x_{(1)} < x_{(2)} < ... < x_{(n)}$ denote the ordered sequence of independent observations from the distribution F, the **empirical distribution function** is defined by

$$\hat{F}(x) = \frac{i}{n+1} \qquad x_{(i)} \leq x < x_{(i+1)}$$

Using this distribution, the **quantile plot** is can be created as the plot of the points

$$\left\{ \left(\hat{F}^{-1} \left(\frac{i}{n+1} \right), x_{(i)} \right) : i = 1, \ldots, n \right\}.$$

Here, x_i and $\hat{F}^{-1}(i/(n+1))$ both gives estimates of the $i/(n+1)$:th quantile of the distribution function F. This means that if $\hat{F}$ is a good estimate of F, the points in the quantile plot should lie close to the unit diagonal.

Using the same conditions as above, the **probability plot** can be created using the points

$$\left\{ \left(\hat{F}^{-1}(x_{(i)}), \frac{i}{n+1} \right) : i = 1, \ldots, n \right\}$$

In the same way as above, the points should be close to the unit diagonal of $\hat{F}$ if the fit is good [Coles, 2001].

2.10.7 Goodness of fit tests for extreme value copulas

To test if a certain copula adequately captures the dependence structure of a dataset, goodness of fit tests can be performed. There are plenty of versions of these tests, this thesis will however restrict to a test for extreme value copulas. If C is a bivariate extreme value copula, it can be proven that

$$C(u, v) = \exp \left[\log(u, v) A \left\{ \frac{\log(v)}{\log(uv)} \right\} \right]$$

for all $u, v \in$ [0,1] and A: [0,1] $\rightarrow$ [1/2,1] representing Pickand's dependence function. A is a convex function that fulfills $Max(1-t, t) \leq A(t) \leq 1$ for all t $\in$ [0,1].

The goodness of fit test compares a parametric estimate A_θ of A with a non-parametric estimate A_n by using Cramér von Mises test statistic

$$S_n = \int_0^1 n|A_n(t) - A_{\theta_n}(t)|^2 \mathrm{d}t \tag{2.8}$$

in order to test the null hypothesis

$$H_0 : A \in \mathcal{A} = \{A_\theta : \theta \in \mathcal{O}\}$$

where $\mathcal{A}$ is a parametric class and $\mathcal{O}$ is an open subset of $\mathbb{R}^d$ for some integer d.

According to Caperaa-Fougeres-Genest a non-parametric estimator is

$$A_n^{CFG}(t) = \exp\left[-\gamma - \frac{1}{n}\sum_{i=1}^{n} \log\{\xi_i(t)\}\right]$$

where $\gamma = \int_0^\infty \log(x)e^{-x}\mathrm{d}x \approx 0.577$ is Euler's constant. $\xi_i(t) = \min(-\log(U_i)/(1-t), -\log(V_i)/t))$ where U_i and V_i are pseudo observations for $i \in \{1 \cdots n\}$.

The test is performed using the following bootstrap algorithm:

1. Compute A_n from the pairs $(U_1, V_1), \cdots, (U_n, V_n)$ and estimate θ using a rank-based estimator.

2. Compute the test statistics defined in Equation (2.8)

3. For some large integer N, repeat the following steps for every $k \in \{1 \cdots N\}$

(a) Generate a random sample $(X_{1k}, Y_{1k}), \cdots, (X_{nk}, Y_{nk})$ from copula C_{θ_n} and calculate the pseudo-observations $(U_{1k}, V_{1k}), \cdots, (U_{nk}, V_{nk})$

(b) Derive A_{nk} and θ_{nk} from $(U_{1k}, V_{1k}), \cdots, (U_{nk}, V_{nk})$

(c) Compute

$$S_{nk} = \int_0^1 n|A_{nk}(t) - A_{\theta_{nk}}|^2 \mathrm{d}t$$

4. An approximate p-value for the test is now given by $\frac{1}{N}\sum_{k=1}^{N(A)} \mathbb{1}(S_{nk} > S_n)$, where $\mathbb{1}$ denotes the indicator function.

[Genest et al., 2011]

Chapter 3

The data

3.1 Choice of data sets

Table (3.1) shows the choices of data series for modelling, being a total of nine. Due to confidentiality reasons the actual times and dates for the data series are not mentioned, but they are all series from the same 1 hour period of the day. The days will furthermore be numbered relative to the first day observed, called "day 1". By choosing data series from different currency pairs, from different days (1 and 2) but also from the same weekday on different weeks (1 and 8) inference might be possible to make on various aspects.

Pair	**day**
EURNOK	1
EURNOK	2
EURNOK	8
EURSEK	1
EURSEK	2
EURSEK	8
EURUSD	1
EURUSD	2
EURUSD	8

Table 3.1 Data sets investigated, all being sampled over a period of 1 hour, being the same for each day. All days are referred to as relative to the first day investigated.

3.2 Form of data

Table (3.2) shows an illustration of the form of the data given, the actual values having been modified due to confidentiality Important properties are the tick data form and the fact that the data at each point in time contains several price levels for bids and offers. For each new point a transaction has been conducted, and a new balance of prices and volumes of bids and offers are given. Note that the actual prices at which the trades occur are not given. As an offer or bid volume can be placed at any price, there are usually volumes for trade present on many prices for both bids and offers at any given point in time. In the data sets available, a maximum of ten price levels containing volumes are given on both the bid- and offer side respectively. The usual scenario being that there are at least ten bid- and offer price levels present at any point in time, there are exceptions where one or both sides contain fewer levels.

Pair	**Time**	**Exchange rates and Bid volumes (Millions)**	**Exchange rates and Offer volumes (Millions)**
EURUSD	01:00:00.124	1.1850 1.1849 ... 1.1842 1.1841 3.000 36.100 ... 23.000 6.000	1.1851 1.1852 ... 1.1859 1.1860 12.900 31.100 ... 3.100 6.000
EURUSD	01:00:00.225	1.1851 1.1850 ... 1.1843 1.1842 4.000 36.300 ... 16.000 9.000	1.1852 1.1851 ... 1.1862 1.1864 10.700 32.100 ... 20.600 12.000
⋮ ⋮	⋮ ⋮	⋮ ⋮	⋮ ⋮
EURUSD	02:00:00.917	1.1820 1.1819 ... 1.1812 1.1811 ... 2.350 30.650 ... 15.000 1.000 ...	1.1822 1.1823 ... 1.1831 1.1832 22.800 20.550 ... 14.500 4.000

Table 3.2 Overview of the example data for one currency pair for illustration purpose, having been modified for confidentiality reasons. In reality the data contains up to 10 price levels at each point in time.

3.3 Data transformations

To enable bivariate modelling of the data, the initial aim is to present volume and price each in the form of univariate data series. As the data at each point in time is given in the form of several price levels with corresponding volumes for bid and offer, there are numerous ways in which to produce univariate representations.

3.3.1 Univariate representations

For the cause of presenting below representations mathematically we introduce the notation which will be used hereafter. Denote at time t the i^{th} offer price level Po_t^i and the corresponding bid price level Pb_t^i, with i denoting the number of price levels from the the mid price of highest big and lowest offer. Using similar notation, the offer- and bid volumes at the same i^{th} level are denoted Vo_t^i and Vb_t^i.

Price representation

The price p_t at time t will be represented simply as the average between the lowest offer and the highest bid price at every point in time, as this is a simple representation with values presumably being close to the actual price levels of the trades taking place. Also, this representation is seemingly unaffected by the volumes, which makes any possible dependence between the two interesting. The representation is

$$p_t = \frac{Po_t^1 + Pb_t^1}{2}.$$

Volume representation

Turning the volume into a univariate data set is not as straight forward. After considering numerous options, all made with the purpose of presenting useful and understandable properties of the data, the measure chosen is the quotient between the total volume present for trade (union of offer and bid volume) and the price span between the highest offer and lowest bid included. This is thus a measure of the concentration of buyers and sellers present in the vicinity of the mid price at each point in time. Letting o_t and b_t denote the number of offer and bid levels respectively containing a volume at time t, with $1 \leq o_t \leq 10$ and $1 \leq b_t \leq 10$, the univariate volume measure v_t is created as

$$v_t = \frac{\sum_{i=1}^{b_t} Vb_t^i + \sum_{i=1}^{o_t} Vo_t^i}{Po_t^{o_t} - Pb_t^{b_t}}$$

3.3.2 Spacing and log-return transformations

While the tick form of the data gives the true picture about the times of transactions, it might complicate modelling and interpretations using time series models.

Spacing

By spacing the data equidistantly in time, data series more suitable for modelling are created. Using linear interpolation, the irregular data might be spaced as preferred.

New updates about the price and volume levels are given at a maximum frequency of 10 Hz. For every update zero or more transactions might have been taken place. The choice of sampling frequency is thus a matter of maximizing the information conserved from the data, by choosing a short distance, but at the same time avoiding too high a frequency which will contain more points between which no transaction has taken place, hence reducing the accuracy of interpolation. The sampling frequency of 10 Hz is chosen, corresponding to a spacing of 0.1 seconds, as this series captures most of the information given and the models derived from this series might be applicable in an appropriate time order.

Log-return

At this point, the price and volume series are univariate and equidistantly spaced in time, see example in the top two plots of Figure (3.1). The log-return transformation is not used on the series, which is a common transformation for time series of price. The simple return, being defined as the quotient between the present and former value, gives a useful measure of price movement from one time to the next. As this quotient is sometimes log-normally distributed, the logarithm transform of the quotient makes it normally distributed and more suitable for modelling. In this thesis work the purpose of the log-return transformation will be rather to produce zero-mean, comparable data, simplifying volatility calculations and further modelling.

Returns at time distance 0.1 and 1 seconds

In order to enable inference to be made at both a short and longer time scale, two types of returns will be created. The first type uses a distance of 0.1 seconds between the points, meaning that every data point in the series is divided by its former value before the logarithm transform. The second type instead uses a distance of 1 seconds between the points, thus only using every 10^{th} value of the series. As this is equivalent to taking the first type log-return on a series spaced with 1 second between the points, this series will furthermore be related to as having "sampling distance 1 second". The two mid plots of Figure (3.1) shows an example of the log-return transformation of price and volume series respectively, using a sampling distance of 1 second.

3.3.3 Calculating volatility

Having produced easily comparable time series, the volatility of each series is finally calculated, as investigation of volatility dependence is the aim of the study. Among several possibilities of representing the volatility σ_t, the Exponentially Weighted Moving Average method explained in Section (2.1), is chosen as it assigns relatively larger weights to recent samples. By implementing this method as a recursive algorithm, the volatility in every point is quickly updated using the latest value. As this method assumes zero-mean values, the null hypothesis of zero mean is first tested for all the series using a simple t-test.

After having calculated the volatility of each data set, the first 500 values are erased from the series sampled at period 0.1 second, and the first 50 values of the series sampled at period 1 second, reason being that the initial volatility values may be unreliable, having been calculated using only a few data points. This is judged to have no significant impact on the results, as the data sets before this removal consist of approximately 36000 and 3600 points respectively, and for illustration purpose, this limit is marked with a red line in Figure (3.1).

The lower two graphs of Figure (3.1) shows the final volatility series of the price and volume of the EURNOK day 1 series, sampled at distance 1 second.

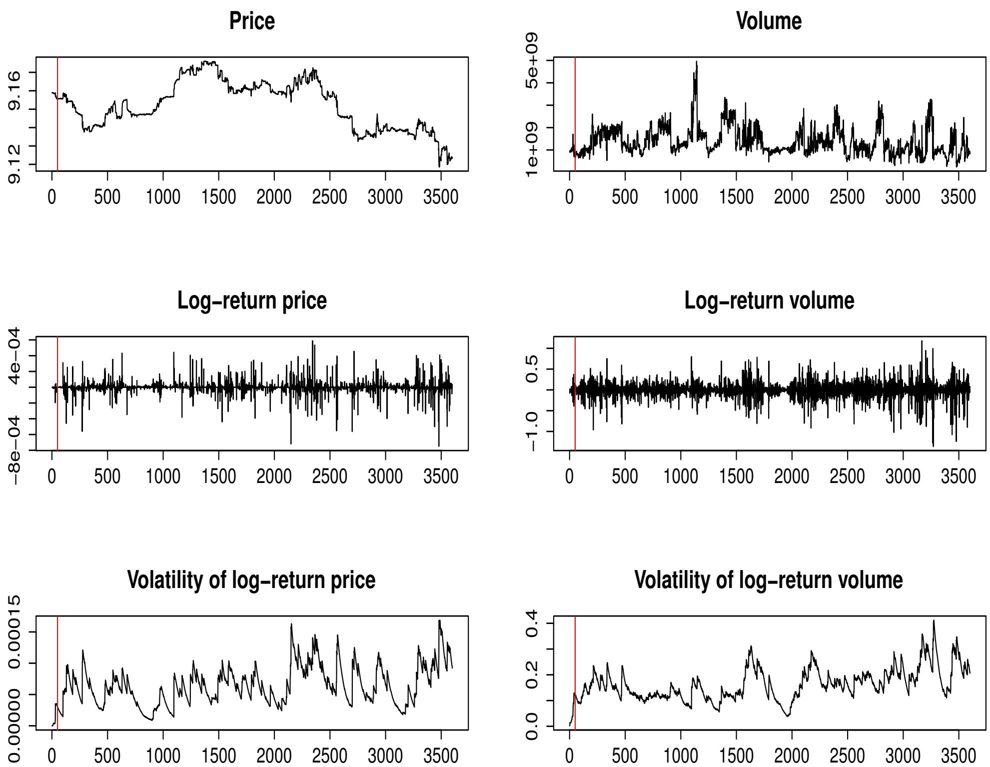


Figure 3.1 Univariate data series (upper), their log-return transformations (mid) and estimated volatilities (lower) for EURNOK day 1, using sampling distance of 1 second. The red line marks the point before which the values are removed due to volatility calculation uncertainty, in this case at 50.

Chapter 4

Analysis and results

In order to carry out extreme value copula modelling of the volatility series, each data set must consist of independent, identically distributed (i.i.d.) variables [Nelson, 2006].

4.1 Time series modelling

One way of removing auto correlations from the data, making it i.i.d., is to fit time series models to the data, through which the data then can be filtered. If fitted correctly, the resulting residuals will possess this quality and thus be suitable for bivariate copula modelling.

The modelling procedure of the time series is conducted in R, using the *forecast* package for ARIMA-modelling and the *rugarch* package if ARCH effects are present in the data, in which case a standard- or exponential GARCH model is added.

ARIMA modelling

As every volatility value was created recursively and thus have a strong correlation at lag 1, the series is first differentiated. This might also account for stochastic trends in the data [Jakobsson, 2013]. The ACF of the series is then plotted. Figure (4.1) shows the EURNOK day 1 differentiated volume volatility series ACF and PACF, the dotted line marking the confidence intervals. As the auto correlation values exceed the confidence intervals at more than 5% of the places, especially for low lags, further ARMA structure is added to the model to complete the ARIMA structure.

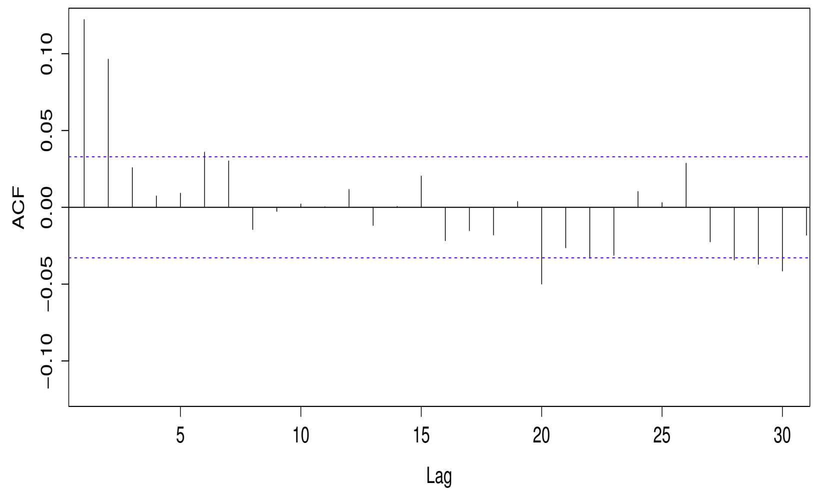


Figure 4.1 ACF of the EURNOK day 1 volume volatility series after differentiating, sampled at distance 1 second.

Due to the large amount of combinations of ARMA parameters possible to fit the differentiated data, the model order is optimized using a stepwise search through the parameter space. The best model is then found as the one minimizing the Akaike Information Criteria (AIC). In order to assure that this is a proper model, model orders in the vicinity of this model is also tested. The residuals of these model are then checked by plotting the ACF of standardized residuals, investigating the weighted Ljung-Box test statistic and the Lagrange Multiplier Test of Engle statistic.

Adding GARCH

In the cases where the Lagrange Multiplier Test of Engle rejects the null hypothesis, showing that ARCH effects are present, a GARCH part is added to the ARIMA model. Even though the two parts account for different effects in the data, the best ARIMA model found eariler is not always the best when combined with a GARCH model. Therefore, several ARIMA models of orders similar to the previously best fitting ARIMA model are tested together with the GARCH model. The standard GARCH is first tried, and the sign bias test statistic is checked for each fitted model. If the null hypothesis of no sign bias is rejected, the exponential GARCH model is instead proposed.

Residual distributions

Before estimating the parameters of an ARIMA-GARCH model, a hypothesis must be made regarding the distribution of the standardized residuals. In order to standardize the residuals, each value in the residual series must be standardized by the standard deviation of its corresponding stochastic variable. Due to this, the residuals of the simple ARIMA model cannot be standardized by a constant standard deviation. If the ARIMA residuals shows ARCH effects, an ARIMA-GARCH model is thus fitted for an assumption of the residual distribution, after which the residuals are standardized by the time varying standard deviation σ_t. A histogram of the residuals is then plotted together with the assumed residual distribution, see example in Figure (4.2).

Due to the very short time between the samples, being 0.1 and 1 seconds respectively, the log-return values and furthermore the volatility estimates are small, giving the standardized residual distributions a high kurtosis. Among the plausible choices of residual distributions, the generalized error distribution (GED) as well as its skewed variant (SGED) is therefore found to be superior to other possibilities such as the normal- or student t distribution. Preceding each final choice of ARIMA-GARCH model, both of these distribution assumptions are tested along with variations of model parameters.

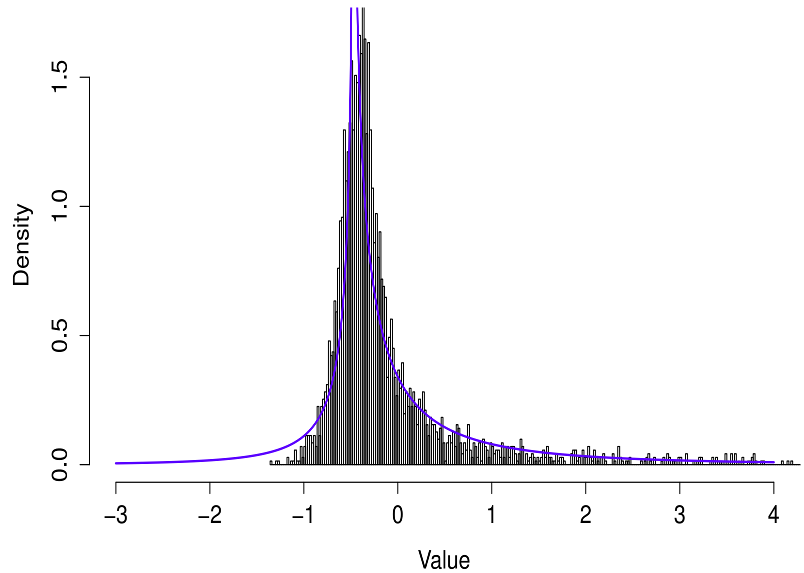


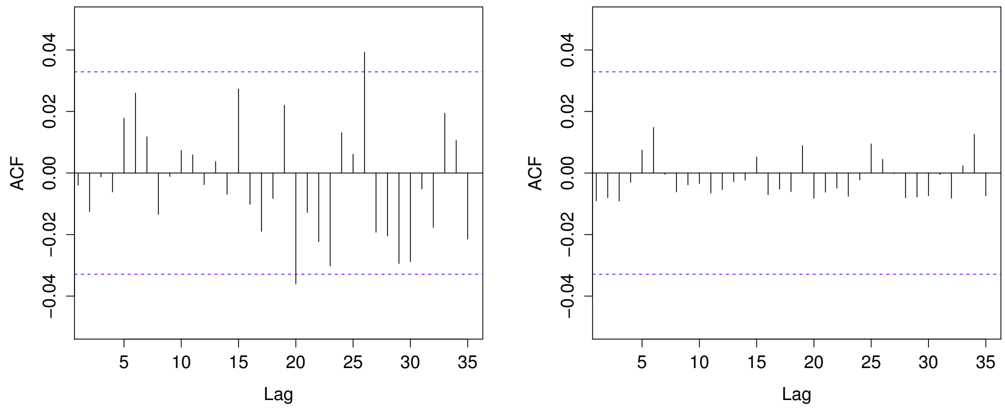
Figure 4.2 Histogram plotted together with assumed skewed generalized error distribution for the standardized residuals, for EURNOK day 1 volume volatility series sampled at distance 1 second.

Reduced data when sampled at 0.1 second

Numerous combinations of model orders and residual distribution assumptions are tested in the process of finding an adequate time series model. Consisting of more than 35000 values, to fit one single model of the series sampled at distance 0.1 is performed in the order of minutes or hours using normal computational power, meaning that finding a suitable model order for one of the 18 models might take time in the magnitude of days. To ensure proper models within a reasonable computation time, the series sampled at 10 Hz are therefore reduced to their initial 10 minutes for further modelling, corresponding to 6000 values.

Final choice of ARIMA-GARCH

The final choice of ARIMA-GARCH model is made using the same whiteness tests as for the pure ARIMA models, assisted also by plotting the ACF of the standardized squared residuals. In the same way that the ACF of the non-squared standardized residuals tests the adequacy of the ARIMA-model, the ACF of the squared standardized residuals tests the adequacy of the GARCH-model. Figure (4.3) illustrates the final ACF plots for the EURNOK day 1 volume volatility series, sampled at distance 1 second. The results for all whiteness tests for sampling distance of 1 and 0.1 seconds are presented in Table (A.3) and (A.4) in the appendix for both fitting of the the price volatility residuals and the volume volatility series. The best models and corresponding parameters can be found in Table (A.1) and (A.2) in the appendix.



(a) ACF of standardized residuals

(b) ACF of squared standardized residuals

Figure 4.3 ACF of standardized and squared standardized residuals from fitted ARMA-GARCH model for EURNOK day 1, indicating and adequately fitted mean- and conditional variance model.

4.2 Bivariate extreme value modelling

Having created white noise residuals, these are now divided into blocks from which the maximum values are extracted. As the residuals are independent, so are also the block maxima, enabling bivariate extreme value copula modelling[Coles, 2001].

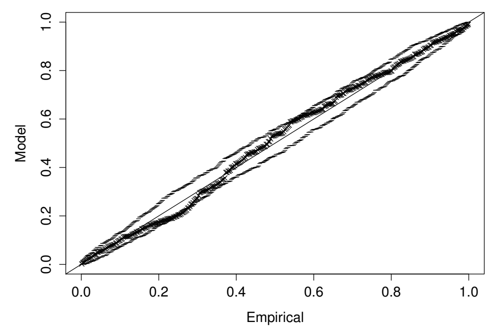
4.2.1 Choosing adequate block sizes

First, a suitable block size must be chosen for modelling. The choice of block size is a trade-off between bias and variance, where too small blocks do not generate block maxima which are adequately approximated by a GEV distribution, whereas too large blocks gives fewer values and thus a larger variance and uncertainty of the estimates [Coles, 2001].

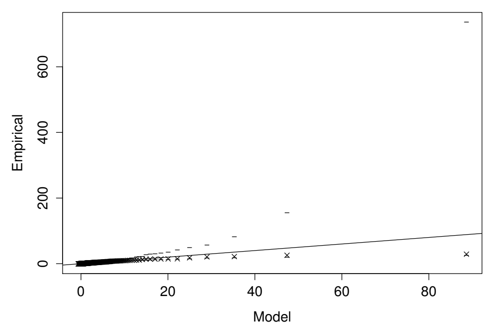
In order to find a common block size suitable for all series of a certain sampling distance, each block maxima series is fitted to a GEV distribution for the given block size and the PP-, QQ- and density plots are investigated to make inference regarding the adequacy of the fit. Figure (4.4) shows these plots for the block maxima of block size 10, for the volume volatility residuals of the EURNOK day 1 series, sampled at distance 1 second. This particular fit barely passes the test, as the PP- and QQ- plot touches the confidence interval points at several occasions.

For the series sampled at distance 1 second, the lowest possible block size producing block maxima adequately fitted by GEV distributions for all series is 10, thus corresponding to a 10 second maxima. This block size is thus chosen for further analysis. A bigger block size of 30 is also tried. As expected, all of these series are also adequately fitted, and this block size is furthermore included for further analysis.

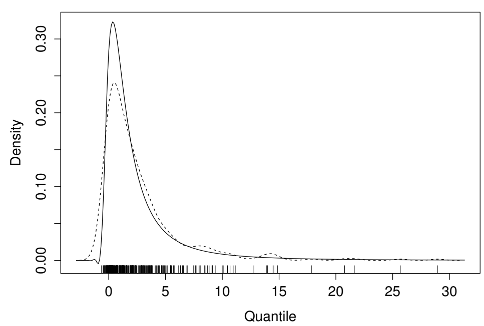
When the same procedure is repeated for the series sampled at distance 0.1 second, the block maxima of both price- and volume volatility residuals are poorly fitted for the data series of EURUSD day 8, EURUSD day 2, EURNOK day 8, EURSEK day 8 for reasonable choices of block size. These are thus excluded for further analysis. For the remaining series, the lowest adequate block size is 50, corresponding to 5 second maxima. The block size of 100 also shows adequate fit, and is included in the analysis. All parameters to the GEV-fitting can be found in Table (B.3) and (B.4).



(a) Probability Plot



(b) Quantile Plot



(c) Density Plot

Figure 4.4 GEV fit diagnostic plots for EURNOK day 1 volume volatility block maxima, showing the PP-, QQ and density plot for block size 10 and sampling distance 1 second.

4.2.2 Dependence structure modelling

In order to evaluate the strength of the dependence between the block maxima of an associated pair of volume- and price volatility residuals, Pearson's correlation coefficient (ρ_p), Kendall's tau (τ) and Spearman's rho (ρ_s) are calculated, to be found in Table (B.1) and (B.2) in the appendix.

An extreme value copula approach is used model their dependence, as not only the dependence strength but also its distribution is of interest. Using the *copula* package of R, the full maximum likelihood (FML) method can be used to evaluate the marginal GEV-parameters and the copula parameter simultaneously, for a copula model of choice. In order to do this, suitable start parameters must be provided to the optimization algorithm. The already fitted GEV distribution parameters provide suitable start parameters for the marginals. In order to provide a suitable copula start parameter value, the extreme value copula models are first fitted to the distribution function values of the marginals, using their fitted GEV distributions. The extreme value copulas that are fitted are those available in the *copula* package, being the Tawn copula, the Husler-Reiss copula, the Galambos copula, the Gumbel copula, the Extremal t copula presented earlier in the theory section. The fitted copula parameters are passed along to the FML optimization together with the marginal GEV parameters. For each FML fit, the goodness of fit of the copula is evaluated, and for the copulas passing this test the one with the lowest AIC value is chosen as the preferred model.

For the best fitted copula model, a 95 % confidence interval is formed both symmetrically, using the standard deviation estimate of the optimization function, and using profile likelihood. The profile likelihood interval is created by first implementing a function which optimizes the marginal parameters by maximum likelihood for a fixed copula parameter. The interval is then created by the two copula parameter values for which this function solves $D_p(\theta_i)$ being equal to the 95 % quantile of the χ_i^2 distribution. Figure 4.5 provides an illustration, corresponding to the best fit of the EURNOK day 1 series, sampled at a distance of 1 second and using block size 10. In this figure, the red lines indicated the profile likelihood interval and the blue lines the symmetric interval, in most cases being very close.

The results for all series can be found in Table (4.1) and (4.2), including best copula found, corresponding goodness of fit p-value, and associated symmetric and profile likelihood confidence intervals. If the lower bound of the confidence interval estimates was found to be negative, it was instead chosen as zero, as the *copula* package parametrization is such that the copula parameters are non-negative.

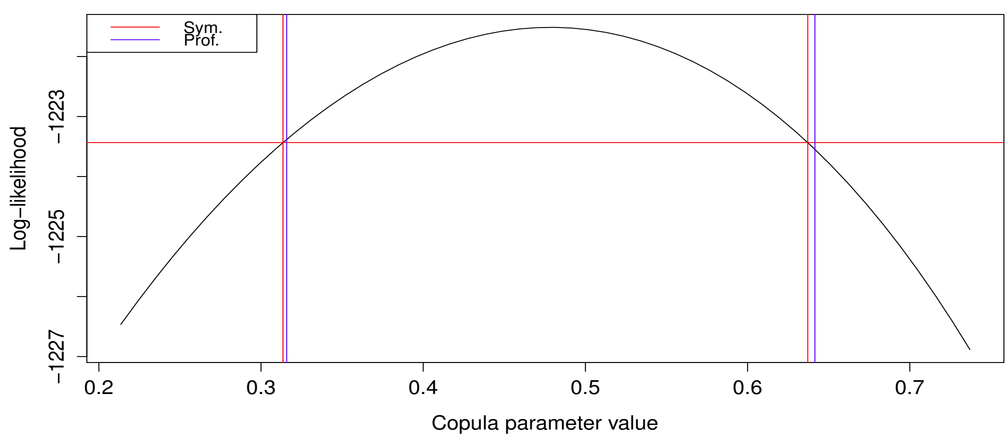


Figure 4.5 Profile likelihood confidence interval marked by vertical red lines, together with corresponding symmetrical interval estimate, marked by blue lines. Example being the EURNOK day 1 series for sampling distance 1 second and block size 10.

Finally, the result from the copula fitting is used to calculate and visually present the conditional probability

$$\mathrm{P}(P > p_x | V = v) = 1 - \mathrm{P}(P \leq p_x | V = v) \tag{4.1}$$

where V and P represents the stochastic variables of price and volatility residual maxima, with corresponding non-stochastic v and p. p_x denotes the x:th quantile of price volatility maxima. Using Bayes' theorem, Equation (4.1) can be reformulated as

$$1 - \mathrm{P}(P \leq p_x | V = v) = 1 - \int_0^{p_x} \frac{f_{V,P}(v,u)}{f_V(v)} \mathrm{d}u$$

where $f_{V,P}(v,p)$ is the joint density function of the variables v and p, and $f_V(v)$ is the marginal density function of variable v. Now, the joint density function can be expressed using copulas according to

$$\begin{aligned} f_{V,P}(v,p) &= \frac{\partial^2}{\partial u \partial v} \; F_{V,P}(v,p) \\ &= \frac{\partial^2}{\partial u \partial v} C(F_V(v), G_P(p)) \\ &= f_V(v) g_P(p) \; c(F_V(v), G_P(p)) \end{aligned}$$

where F_V and G_P denotes the marginal distribution functions for V and P respectively, and f_V and g_P their corresponding density functions.

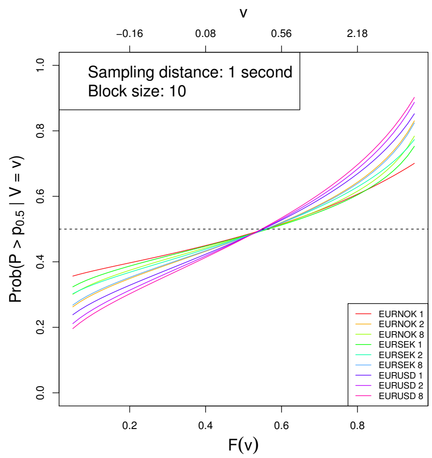
In this work, the probability is calculated for a fixed value of of p_x and various values of v, then plotted as a function of v and $F(v)$, using a joint density function defined via the estimated best copula. The result can be seen in Figure (4.6) for sampling distance 1 and corresponding block sizes. Figure (4.7) shows the corresponding plots for sampling distance 0.1 second. For both sampling distances, p_x is calculated for $x = 0.5$ and $x = 0.8$, corresponding to the 50% and 80% quantiles respectively. In the case of no correlation between the variables, the $p_{0.8}$ quantile should be exceeded with a probability of 20%, which is marked by a horizontal dashed line.

Currency	Date	Copula	θ	GoF	Symmetric C.I	Profile C.I
Sampling distance: 1 second. Block size: 10						
EURUSD	1	Galambos	0.727	0.255	[0.585, 0.869]	[0.592, 0.879]
EURUSD	2	Husler-Reiss	1.229	0.053	[1.058, 1.399]	[1.066, 1.408]
EURUSD	8	Husler-Reiss	1.282	0.478	[1.107, 1.458]	[1.114, 1.467]
EURSEK	1	Galambos	0.529	0.988	[0.416, 0.641]	[0.422, 0.648]
EURSEK	2	Gumbel	1.314	0.984	[1.195, 1.433]	[1.204, 1.442]
EURSEK	8	Husler-Reiss	1.052	0.146	[0.901, 1.203]	[0.907, 1.210]
EURNOK	1	Tawn	0.479	0.490	[0.316, 0.641]	[0.314, 0.637]
EURNOK	2	Husler-Reiss	1.068	0.263	[0.916, 1.220]	[0.922, 1.227]
EURNOK	8	Husler-Reiss	0.961	0.858	[0.817, 1.106]	[0.814, 1.104]
Sampling distance: 1 second. Block size: 30						
EURUSD	1	Husler-Reiss	1.085	0.189	[0.803, 1.366]	[0.822, 1.392]
EURUSD	2	Husler-Reiss	0.857	0.253	[0.572, 1.143]	[0.562, 1.156]
EURUSD	8	Husler-Reiss	1.000	0.172	[0.741, 1.259]	[0.757, 1.282]
EURSEK	1	Husler-Reiss	1.001	0.216	[0.733, 1.270]	[0.746, 1.292]
EURSEK	2	Tawn	0.461	0.547	[0.198, 0.724]	[0.197, 0.712]
EURSEK	8	Husler-Reiss	1.213	0.145	[0.867, 1.559]	[0.891, 1.592]
EURNOK	1	Husler-Reiss	0.796	0.326	[0.547, 1.046]	[0.531, 1.057]
EURNOK	2	Galambos	0.588	0.781	[0.376, 0.802]	[0.393, 0.826]
EURNOK	8	Husler-Reiss	0.852	0.848	[0.585, 1.119]	[0.583, 1.134]

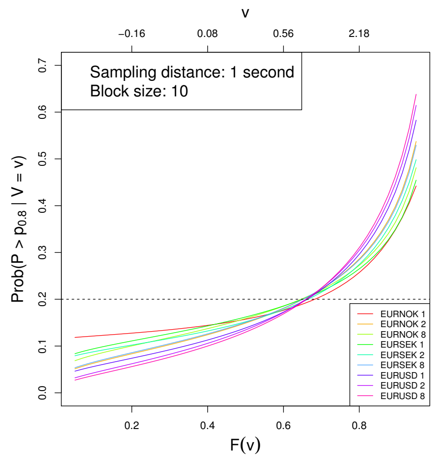
Table 4.1 Copula results of all series tested for sampling distance 1 second, including type of copula, the copula parameter θ, the corresponding goodness of fit p-value, and the symmetric and profile likelihood confidence intervals. In the case where the lower limit of the symmetric- or profile interval is negative, this has been fixed to 0 as no negative values are possible.

Currency	Date	Copula	$\boldsymbol{\theta}$	GoF	Symmetric C.I	Profile C.I
Sampling distance: 0.1 second. Block size: 50						
EURUSD	1	Tawn	0.225	0.363	[0.000, 0.542]	[0.000, 0.538]
EURUSD	2	-	-	-	-	-
EURUSD	8	-	-	-	-	-
EURSEK	1	Gumbel	1.319	0.917	[1.117, 1.521]	[1.141, 1.549]
EURSEK	2	Galambos	0.315	0.386	[0.087, 0.544]	[0.000, 0.536]
EURSEK	8	-	-	-	-	-
EURNOK	1	Tawn	0.224	0.845	[0.000, 0.548]	[0.000, 0.533]
EURNOK	2	Tawn	0.445	0.581	[0.161, 0.729]	[0.157, 0.713]
EURNOK	8	-	-	-	-	-
Sampling distance: 0.1 second. Block size: 100						
EURUSD	1	Gumbel	1.020	0.385	[0.856, 1.183]	[0.000, 1.229]
EURUSD	2	-	-	-	-	-
EURUSD	8	-	-	-	-	-
EURSEK	1	Husler-Reiss	0.709	0.640	[0.376, 1.042]	[0.000, 1.059]
EURSEK	2	Tawn	0.212	0.445	[0.000, 0.649]	[0.000, 0.637]
EURSEK	8	-	-	-	-	-
EURNOK	1	Tawn	0.038	0.126	[0.000, 0.554]	[0.000, 0.521]
EURNOK	2	Husler-Reiss	0.907	0.430	[0.569, 1.244]	[0.590, 1.288]
EURNOK	8	-	-	-	-	-

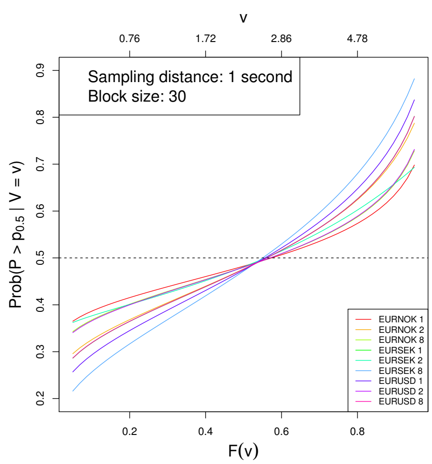
Table 4.2 Copula results of all series tested for sampling distance 0.1 second, including type of copula, the copula parameter θ, the corresponding goodness of fit p-value, and the symmetric and profile likelihood confidence intervals. In the case where the lower limit of the symmetric- or profile interval is negative, this has been fixed to 0 as no negative values are possible.



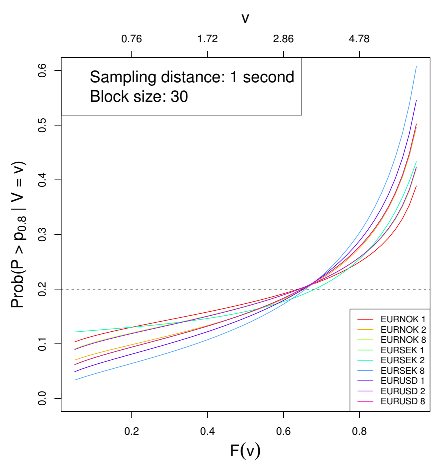
(a) Block size 10 and p is the 0.5 quantile.



(b) Block size 10 and p is the 0.8 quantile.

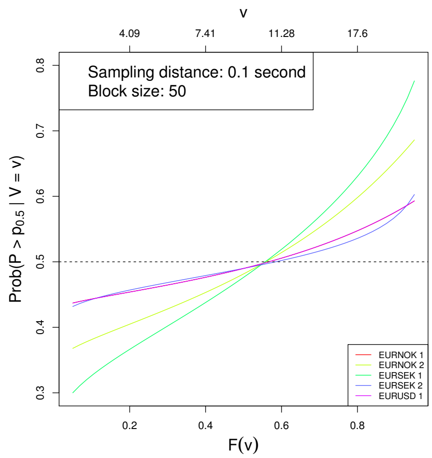


(c) Block size 30 and p is the 0.5 quantile.

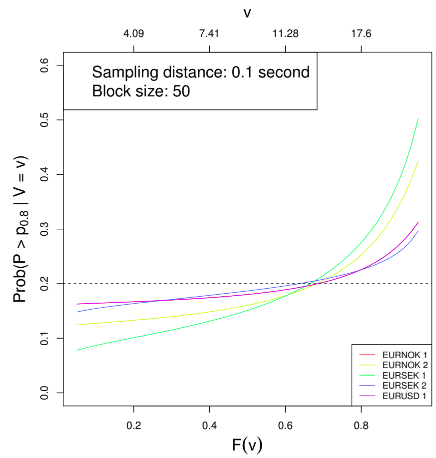


(d) Block size 30 and p is the 0.8 quantile.

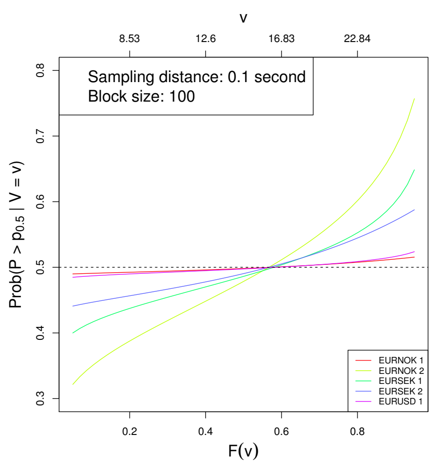
Figure 4.6 The conditional probability $\mathrm{P}(P > p_x | V = v)$ plotted against v for the currency pairs used for sampling distance 1 second. Used quantile probabilities are $x = 0.5$ and 0.8. The block size is 10 for the upper figures and 30 for the lower.



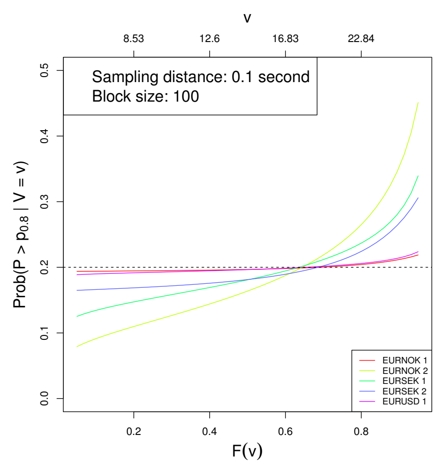
(a) Block size 50 and p is the 0.5 quantile.



(b) Block size 50 and p is the 0.8 quantile.



(c) Block size 100 and p is the 0.5 quantile.



(d) Block size 100 and p is the 0.8 quantile.

Figure 4.7 The conditional probability $\mathrm{P}(P > p_x | V = v)$ plotted against v for the currency pairs used for sampling distance 0.1 second. Used quantile probabilities are $x = 0.5$ and 0.8. The block size is 50 for the upper figures and 100 for the lower.

Chapter 5

Conclusions and recommendations

5.1 Conclusions

All data series having a sampling distance of 1 second could successfully be fitted to a copula model, as well as all series with a sampling distance of 0.1 second having block maxima adequately fitted by a GEV distribution. As this is the case regardless of the block size and sampling distance investigated, it provides strong evidence for the hypothesis that there is dependence between volume- and price volatility maxima. The confidence interval of the copula parameter is in most cases similar for the symmetric normal approximation and for the interval constructed using profile likelihood, being significantly positive for all series sampled at distance 1. The results for sampling distance 0.1 is weaker and in around half of the cases inconclusive as the confidence intervals reach the zero.

The hypothesis of significant dependence existing between volume- and price volatility extreme values can be considered verified, though seemingly dependent on e.g. sampling distance.

Regarding which copula type that best models the extreme dependence, a variety of conclusions can be drawn from Table (4.1) and (4.2). For sampling distance of 1 second, the Husler-Reiss copula is generally found to provide the best fit, accounting for 12 of the 18 optimal copulas dependence structures being evaluated, also being the generally best fit within each block size specifically. For sampling distance of 0.1 second, the Tawn copula is the slightly doubtful winner accounting for about half of the best fits for each block size. A specifically clear result is that of sampling distance 1 second and block size 30, for which all series of day 1 and 8, having the same weekday, are best fitted by the Husler-Reiss copula.

Considering above conclusions, it should be kept in mind that other copulas than the best might also provide good fits.

Assessing the strength of the dependence, Table (B.1) and (B.2) show similarities in Kendall's τ across block sizes but within the same sampling distance. The values are in the region of 0.1 - 0.3 for sampling distance of 1 second and slightly lower for sampling distance 0.1 second, which provides evidence that the sampling distance has an effect on the strength of the dependence. The 0.1 second results not only show weaker dependence in general, but it is also here that the confidence intervals in some cases reach zero, making the results partly inconclusive. This is especially the case for block size 100, where only the EURNOK 2 series is found to have a significant dependence between the maxima judging by the profile intervals.

The conditional probability plots of Figure (4.6) and (4.7) strengthen above conclusions, also providing further information about the results. Although being a simple deduction from the copula model and fitted parameter, these plots visualize the results clearly. Not including confidence intervals, they should though be analyzed with caution, particularly for the results of sampling distance 0.1 second.

As compared to the dashed line, indicating no dependence, the positive slope of all curves visualize the positive dependence. The tail dependence of the estimates relates to the vertical distance from the dashed line to the probability estimate, for low and high volume extremes respectively. The slope of the curve reveals for which values of the volume volatility that the dependence is increasing the most. The increase is most prominent in the upper tail of all series. The strongest upper tail dependence is found for the series sampled at a 1 second distance, most importantly visualized in the graphs in Figure (4.6), using the 0.8 quantile. This is also confirmed by the scatter plots of in Figure (B.1) and (B.2). Focusing on the block size 10 plots, an interesting observation is that the three EURUSD series have the strongest tail dependence of all series, though this conclusion cannot be generalized to block size 30. Assessing the results of sampling distance 0.1, the prominent results are those of EURNOK 2 and EURSEK 1, showing the strongest overall dependence regardless of block size, also having significantly non-zero copula parameter values as shown in Table (4.1) and (4.2).

All graphs of each plot intersect at the same point, which in addition nearly coincides with the dashed line. The value of v where this occurs seems to be independent on block size, but dependent on quantile p_x. This common point marks a value of the volume volatility residual extreme v for which all series show independence between volume and price volatility maxima. Volume volatility extremes exceeding value will likely be associated with larger price volatility extremes than in the independent case. Smaller volume volatility extremes will on the contrary be associated with smaller price volatility extremes than in the independent case.

5.2 Method and errors

The thesis work has to some extent included previously unknown concepts, such as the representation of volume "concentration" and "volatility", followed by several steps of data transformations resulting in price and volume volatility series. To name an example, the choice of weight λ for the volatility calculations was rather subjective, since no standard value for high frequency data was found.

Modelling the time series to remove auto correlations, the model parameters were assumed to be constant, which seemed to be a fair assumption since the total length of the time series amounted to no more than one hour. However, allowing time varying parameters might have resulted in better time series models. The fitting of the series sampled at distance 1 second proved easier than the 0.1 second sequences. A predominant cause of this is presumably that the higher the sampling frequency, the smaller the change between subsequent values, which makes the log-return values close to or equal zero, resulting in small volatility values. Judging by the very high kurtosis of the standardized residuals when sampling at a 1 second distance, example provided in Figure (4.2), the 0.1 second distributions proved to have an even larger kurtosis. Even so, it was possible to find suitable time series models for both sampling frequencies judging by the residual whiteness tests results in Table (A.3) and (A.4).

Finding a suitable, common block size for the residual data proved hard, especially for the 0.1 second series. Only by excluding the four generally ill-fitting 0.1 second sets this was made possible. Also, even though the diagnostic plots showed adequate fits of all series, individual choices of block sized would have allowed for even better fits, at the cost of less generalized results and conclusions.

By trying five different extreme value copula models to model the dependence, choosing the best model passing the goodness of fit test by utilizing Akaikes Information Criteria, and finally optimizing the marginal and copula parameters simultaneously, the final results of Table (4.1) and (4.2) are considered valid. The similarity between the symmetric and profile intervals, in most cases being significantly positive, further strengthens the validity of the results. The information criteria values founding the choice of best copula model were in many cases similar and other copula models than the chosen one might have adequately fitted the data, enabling further depth of conclusions.

The calculations underlying the conditional probability plots might be prone to errors, due to the numerical methods used to calculate each point. Including several steps of multiplication of copula- and marginal functions, finalized by numerical integration, small errors in each model might build up to larger errors in the final graphs. Thus these should be regarded with caution while lacking confidence intervals.

5.3 Recommendations

This thesis research has focused on an unexplored topic, using the quite novel technique of extreme value copulas. Without preceding suggestions that there would be dependence to be found, the results show evidence for dependence between price and volume volatility above expectations. We therefore suggest further work on the topic.

To test the conclusions of this work, the models found should first be verified on more data, producing a more holistic and precise picture of the dependence. For instance the Husler-Reiss copula, seeming to be the best model for sampling distance 1 second, could be examined in further detail on all current data sets but also additional data. By also allowing time varying time series- and extreme value models throughout all hours of trade, conclusion strength would be improved.

Finally, the foremost topic of further analysis would be to investigate the exact causation between price- and volume volatility, including both extremes and non-extremes. Having evaluated this dependence structure, the results might be applicable enough to include as extensions to existing trading models.

Appendix A

Time series modelling

A.1 Time series models

Currency pair	Date	a_1	a_2	a_3	d	c_1	c_2	c_3	c_4	ω	α_1	β_1	γ	skew	shape
Volume volatility models, sampling distance 1 second.															
EURUSD	1	-	-	-	1	0.159	0.071	-	-	-	-	-	-	-	-
EURUSD	2	0.830	-0.990	-	1	-0.663	0.934	0.103	0.090	-	-	-	-	-	-
EURUSD	8	0.136	-	-	1	-	-	-	-	-	-	-	-	-	-
EURSEK	1	-	-	-	1	0.411	-	-	-	-2.367	-0.040	0.766	0.476	-	0.785
EURSEK	2	-0.984	0.025	-	1	-0.885	-	-	-	0.000	0.152	0.847	-	2.223	0.523
EURSEK	8	-0.029	0.010	-	1	0.393	-	-	-	-	-	-	-	-	-
EURNOK	1	-0.237	-0.129	-	1	-	-	-	-	0.000	0.684	0.297	-	1.890	0.517
EURNOK	2	-0.981	0.064	-0.015	1	-0.826	-	-	-	0.000	0.3014	0.698	-	2.206	0.476
EURNOK	8	-	-	-	1	0.343	-	-	-	-1.022	-0.120	0.909	0.294	-	0.682
Price volatility models, sampling distance 1 second															
EURUSD	1	0.093	0.035	0.059	1	-	-	-	-	-	-	-	-	-	-
EURUSD	2	-	-	-	1	-	-	-	-	-	-	-	-	-	
EURUSD	8	-	-	-	1	0.059	-	-	-	-	-	-	-		
EURSEK	1	0.410	-	-	1	-0.318	-	-	-	-	-	-	-		
EURSEK	2	-0.985	-	-	1	-0.022	-	-	-	-4.389	0.959	0.846	0.099	-	0.621
EURSEK	8	-	-	-	1	-	-	-	-	-0.031	-0.071	0.988	0.127	-	0.100
EURNOK	1	0.618	-	-	1	0.556	-	-	-	-	-	-	-	-	-
EURNOK	2	-	-	-	1	0.147	-	-	-	-	-	-	-	-	-
EURNOK	8	-	-	-	1	0.221	-	-	-	-0.085	-0.116	0.997	0.176	1.426	0.333

Table A.1 Time series model parameters for the volume volatility, using a sampling distance of 1 second.

Currency pair	Date	a_1	a_2	a_3	d	c_1	c_2	c_3	c_4	ω	α_1	β_1	γ	skew	shape
Volume volatility models, sampling distance 0.1 seconds															
EURUSD	1	0.333	0.645	-0.003	1	0.641	-0.009	0.001	-	-2.326	0.612	0.825	0.112	-	0.190
EURUSD	2	0.980	0.005	-	1	-	-	-	-	0.000	0.385	0.563	-	1.000	0.154
EURUSD	8	0.782	0.200	-	1	0.203	-	-	-	0.000	1.000	0.728	-	-	0.100
EURSEK	1	-	-	-	1	0.720	0.349	-	-	0.000	0.640	0.359	-	-	0.595
EURSEK	2	0.985	-	-	1	-	-	-	-	0.000	0.321	0.670	-	1.000	0.158
EURSEK	8	0.985	-	-	1	-	-	-	-	0.000	0.321	0.670	-	1.000	0.158
EURNOK	1	0.237	0.737	-	1	0.823	0.217	0.121	-	0.000	1.000	0.602	-	-	0.254
EURNOK	2	0.803	0.179	-	1	-0.489	-0.196	-0.058	-	0.000	1.000	0.526	-	1.827	0.374
EURNOK	8	0.771	0.211	-	1	0.214	-	-	-	0.000	1.000	0.464	-	-	0.100
Price volatility models, sampling distance 0.1 seconds															
EURUSD	1	0.985	-	-	1	-0.010	-	-	-	-8.212	1.316	0.749	0.097	-	0.100
EURUSD	2	0.985	-	-	1	-	-	-	-	-8.062	-0.072	0.890	0.133	-	0.106
EURUSD	8	0.514	0.407	0.056	1	0.412	-	-	-	-2.814	-0.112	0.935	0.115	1.000	0.124
EURSEK	1	0.993	-	-	1	-0.004	-	-	-	-3.026	-0.037	0.914	0.101	-	0.172
EURSEK	2	-	-	-	1	0.832	0.264	-	-	-0.203	-0.078	0.994	0.112	1.183	0.341
EURSEK	8	-	-	-	1	0.601	-	-	-	-0.318	-0.170	0.990	0.190	-	0.411
EURNOK	1	0.985	-	-	1	-0.002	-	-	-	-8.279	-0.060	0.775	0.143	-	0.100
EURNOK	2	-	-	-	1	0.373	0.132	-	-	-0.046	-0.159	1.000	0.190	1.733	0.297
EURNOK	8	0.994	-	-	1	-	-	-	-	0.000	0.050	0.899	-	-	0.225

Table A.2 Time series model parameters for the volume volatility, using a sampling distance of 0.1 second.

A.2 Whiteness tests

Currency pair	Date	W-LB ST	W-LB SQ	ARCH LM	Sign bias
Volume residual series, sampling distance 1 seconds					
EURUSD	1	0.960	0.799	1.000	-
EURUSD	2	0.805	0.000	0.078	-
EURUSD	8	0.917	0.033	0.997	-
EURSEK	1	0.030	0.756	1.000	0.563
EURSEK	2	0.856	0.938	1.000	0.528
EURSEK	8	0.986	0.036	0.967	-
EURNOK	1	0.812	0.590	0.999	0.715
EURNOK	2	0.212	0.844	0.828	0.040
EURNOK	8	0.052	0.950	1	0.835
Price residual series, sampling distance 1 seconds					
EURUSD	1	0.918	0.132	0.997	-
EURUSD	2	0.085	0.986	1.000	-
EURUSD	8	0.970	0.302	0.998	-
EURSEK	1	0.963	0.352	1.000	-
EURSEK	2	0.227	0.933	1.000	0.451
EURSEK	8	0.528	0.931	1.000	0.103
EURNOK	1	0.990	0.889	0.940	-
EURNOK	2	0.861	0.085	0.989	-
EURNOK	8	0.672	0.981	1.000	0.145

Table A.3 Results of Weighted Ljung-Box test of standardized and squared standardized residuals, ARCH LM test and sign bias for the ARIMA-GARCH models. The sampling distance is 1 second. Note that the sign bias test is only performed for models having a GARCH-part.

Currency pair	Date	W-LB ST	W-LB SQ	ARCH LM	Sign bias
Volume residual series, sampling distance 0.1 seconds					
EURUSD	1	0.292	0.148	0.512	0.551
EURUSD	2	0.387	0.944	0.941	0.173
EURUSD	8	0.899	0.975	0.975	0.058
EURSEK	1	0.544	0.936	0.905	0.665
EURSEK	2	0.725	0.951	0.951	0.672
EURSEK	8	0.869	0.976	0.976	0.564
EURNOK	1	0.734	0.978	0.979	0.268
EURNOK	2	0.963	1.000	1.000	0.831
EURNOK	8	0.957	0.978	0.978	0.348
Price residual series, sampling distance 0.1 seconds					
EURUSD	1	0.557	0.985	0.985	0.704
EURUSD	2	0.676	0.997	0.992	0.478
EURUSD	8	0.923	0.897	0.891	0.699
EURSEK	1	0.548	0.972	0.970	0.549
EURSEK	2	0.248	0.989	0.985	0.890
EURSEK	8	0.969	0.982	0.982	0.650
EURNOK	1	0.803	0.975	0.975	0.607
EURNOK	2	0.492	0.986	0.981	0.870
EURNOK	8	0.568	0.984	0.981	0.916

Table A.4 Results of Weighted Ljung-Box test of standardized and squared standardized residuals, ARCH LM test and sign bias for the ARIMA-GARCH models. The sampling distance is 0.1 second. Note that the sign bias test is only performed for models having a GARCH-part.

Appendix B

Block maxima modelling

B.1 BM dependence measures

Currency	Date	Pearson's ρ	Spearman's ρ	Kendall's τ
Sampling distance: 1 second. Block size: 10				
EURNOK	1	0.331	0.278	0.193
EURNOK	2	0.496	0.401	0.274
EURNOK	8	0.356	0.342	0.229
EURSEK	1	0.331	0.343	0.235
EURSEK	2	0.434	0.301	0.204
EURSEK	8	0.130	0.445	0.308
EURUSD	1	0.703	0.426	0.297
EURUSD	2	0.309	0.508	0.348
EURUSD	8	0.534	0.478	0.326
Sampling distance: 1 second. Block size: 30				
EURUSD	1	0.803	0.400	0.281
EURUSD	2	0.207	0.281	0.193
EURUSD	8	0.538	0.354	0.245
EURSEK	1	0.322	0.349	0.238
EURSEK	2	0.395	0.178	0.124
EURSEK	8	0.052	0.411	0.276
EURNOK	1	0.300	0.191	0.130
EURNOK	2	0.526	0.309	0.215
EURNOK	8	0.474	0.253	0.167

Table B.1 Dependence measures between block maxima of residual series sampled at distance 1 second.

Currency	**Date**	**Pearson's ρ**	**Spearman's ρ**	**Kendall's τ**
Sampling distance: 0.1 second. Block size: 50				
EURNOK	1	-0.058	0.088	0.054
EURNOK	2	0.104	0.200	0.136
EURNOK	8	-	-	-
EURSEK	1	0.199	0.320	0.220
EURSEK	2	-0.020	0.135	0.090
EURSEK	8	-	-	-
EURUSD	1	0.026	0.104	0.066
EURUSD	2	-	-	-
EURUSD	8	-	-	-
Sampling distance: 0.1 second. Block size: 100				
EURNOK	1	-0.076	0.087	0.047
EURNOK	2	0.736	0.254	0.173
EURNOK	8	-	-	-
EURSEK	1	0.120	0.234	0.130
EURSEK	2	-0.002	0.152	0.099
EURSEK	8	-	-	-
EURUSD	1	-0.035	0.034	0.012
EURUSD	2	-	-	-
EURUSD	8	-	-	-

Table B.2 Dependence measures between block maxima of residual series sampled at distance 0.1 second.

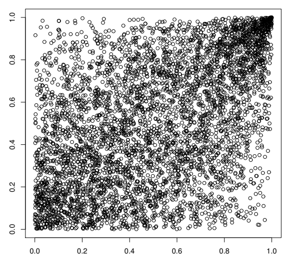
B.2 GEV Models

Currency	**Date**	**loc V**	**scale V**	**shape V**	**loc P**	**scale P**	**shape P**
Sampling distance: 1 second. Block size: 10							
EURUSD	1	0.422	0778	0.517	0.626	0.829	0.403
EURUSD	2	0.178	0.513	0.810	0.341	0.876	0.719
EURUSD	8	0.030	0.520	1.183	0.368	0.779	0.711
EURSEK	1	0.612	1.159	0.561	0.070	0.399	0.700
EURSEK	2	0.839	1.471	0.651	0.373	0.631	1.400
EURSEK	8	0.175	0.483	0.672	0.018	0.671	1.236
EURNOK	1	0.868	1.303	1.303	0.072	0.457	0.912
EURNOK	2	0.825	1.355	0.698	0.162	0.581	0.746
EURNOK	8	0.575	1.066	0.636	-0.128	0.276	1.732
Sampling distance: 1 second. Block size: 30							
EURUSD	1	1.624	1.275	0.239	1.704	1.194	0.260
EURUSD	2	0.919	1.193	0.733	2.158	1.788	0.149
EURUSD	8	1.562	1.787	0.236	1.604	1.487	0.365
EURSEK	1	2.455	2.361	0.313	0.482	0.799	1.109
EURSEK	2	3.754	2.848	0.318	2.323	2.482	0.708
EURSEK	8	0.753	1.010	0.928	1.716	2.814	1.034
EURNOK	1	2.867	2.629	0.368	1.099	1.485	0.731
EURNOK	2	3.702	3.046	0.419	1.100	1.401	0.632
EURNOK	8	2.423	2.358	0.497	0.661	1.699	1.689

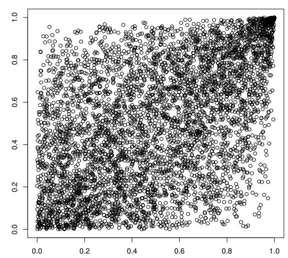
Table B.3 GEV fit parameters for the volume (V) volatility residual extremes and price (P) volatility residual extremes. FML method has been used to fit the GEV distributions. Sampling distance is 1 second.

Currency	Date	loc V	scale V	shape V	loc P	scale P	shape P
Sampling distance: 0.1 second. Block size: 50							
EURUSD	1	6.870	6.127	0.202	13.839	10.763	0.332
EURUSD	2	-	-	-	-	-	-
EURUSD	8	-	-	-	-	-	-
EURSEK	1	3.396	4.247	0.837	8.557	12.453	0.972
EURSEK	2	15.398	17.147	0.912	0.435	1.421	2.135
EURSEK	8	-	-	-	-	-	-
EURNOK	1	10.662	10.681	1.128	34.825	53.461	1.176
EURNOK	2	12.534	13.759	0.981	2.888	4.390	0.971
EURNOK	8	-	-	-	-	-	-
Sampling distance: 0.1 second. Block size: 100							
EURUSD	1	11.964	7.230	0.004	21.938	12.848	0.428
EURUSD	2	-	-	-	-	-	-
EURUSD	8	-	-	-	-	-	-
EURSEK	1	7.754	7.730	0.660	26.511	26.936	0.589
EURSEK	2	24.500	24.640	1.230	4.256	6.505	0.979
EURSEK	8	-	-	-	-	-	-
EURNOK	1	22.611	21.923	1.025	101.504	121.089	0.825
EURNOK	2	23.259	24.107	1.086	9.920	10.496	0.553
EURNOK	8	-	-	-	-	-	-

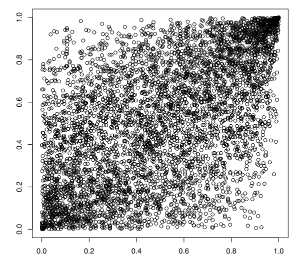
Table B.4 GEV fit parameters for the volume (V) volatility residual extremes and price (P) volatility residual extremes. FML method has been used to fit the GEV distributions. Sampling distance is 0.1 second.



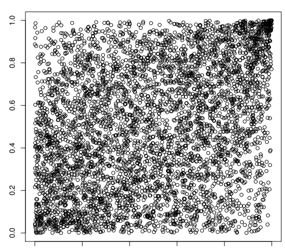
(a) EURUSD 1
Galambos



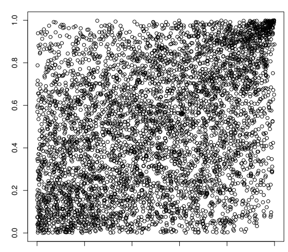
(b) EURUSD 2
Husler-Reiss



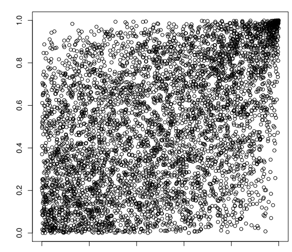
(c) EURUSD 8
Husler-Reiss



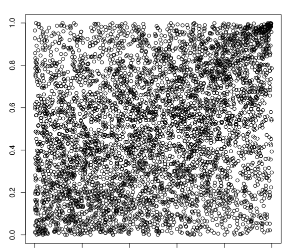
(d) EURSEK 1
Galambos



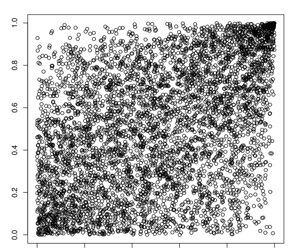
(e) EURSEK 2
Gumbel



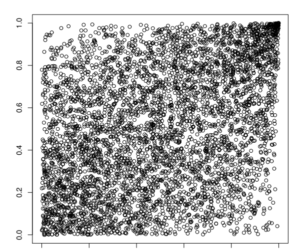
(f) EURSEK 8
Husler-Reiss



(g) EURNOK 1
Tawn

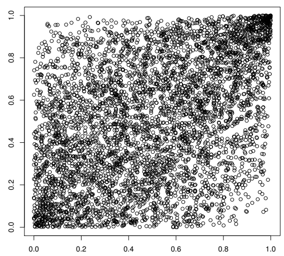


(h) EURNOK 2
Husler-Reiss

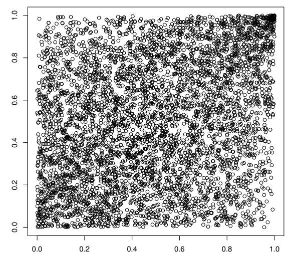


(i) EURNOK 8
Husler-Reiss

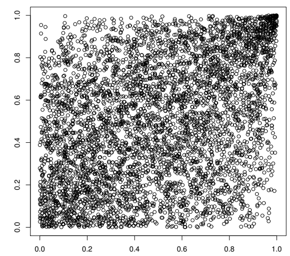
Figure B.1 Scatter plots of randomizations from each best copula. Above plots corresponds to block size 10 and sampling distance 1 second.



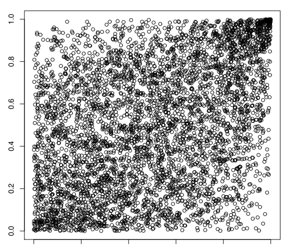
(a) EURUSD 1
Husler-Reiss



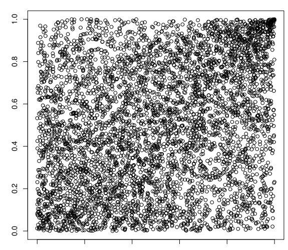
(b) EURUSD 2
Husler-Reiss



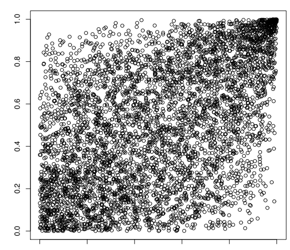
(c) EURUSD 8
Husler-Reiss



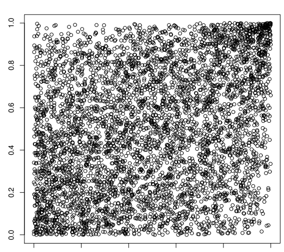
(d) EURSEK 1
Husler-Reiss



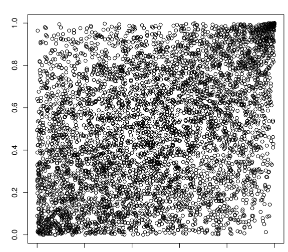
(e) EURSEK 2
Tawn



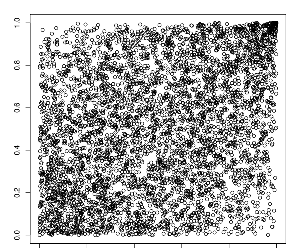
(f) EURSEK 8
Husler-Reiss



(g) EURNOK 1
Husler-Reiss

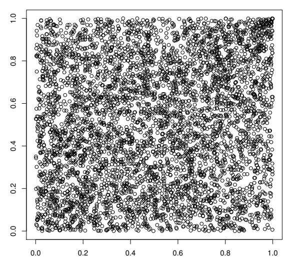


(h) EURNOK 2
Galambos

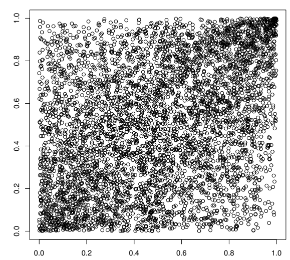


(i) EURNOK 8
Husler-Reiss

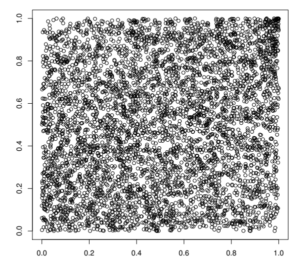
Figure B.2 Scatter plots of randomizations from each best copula. Above plots corresponds to block size 30 and sampling distance 1 second.



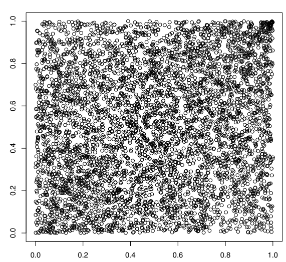
(a) EURUSD 1
Tawn



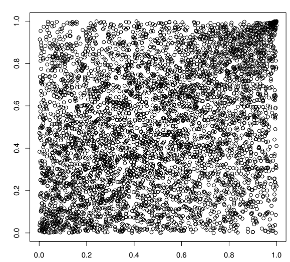
(b) EURSEK 1
Gumbel



(c) EURSEK 2
Galambos

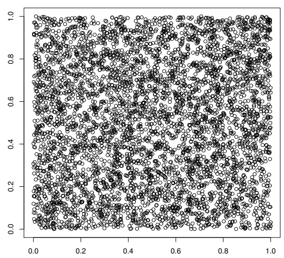


(d) EURNOK 1
Tawn

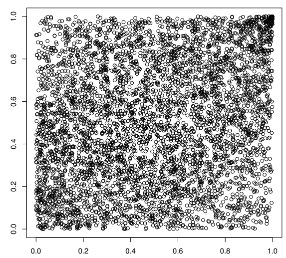


(e) EURNOK 2
Tawn

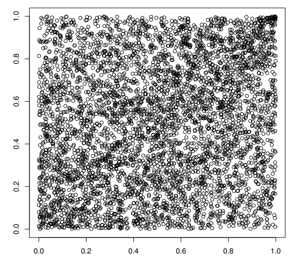
Figure B.3 Scatter plots of randomizations from each best copula. Above plots corresponds to block size 50 and sampling distance 0.1 second.



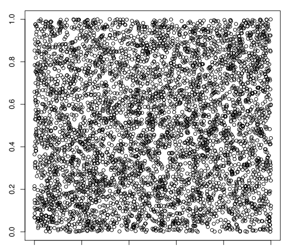
(a) EURUSD 1
Gumbel



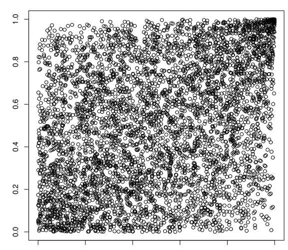
(b) EURSEK 1
Husler-Reiss



(c) EURSEK 2
Tawn



(d) EURNOK 1
Tawn



(e) EURNOK 2
Husler-Reiss

Figure B.4 Scatter plots of randomizations from each best copula. Above plots corresponds to block size 100 and sampling distance 0.1 second.

Bibliography

Irene Aldridge. *High Frequency Trading.* John Wiley and Sons, Inc., Hoboken, New Jersey, 2010.

Stuart Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer, 2001.

Stefano Demarta and Alexander J. McNeil. Extreme value copulas and max-stable processes. *Journal de la Société Française de Statistique*, 2005.

Kevin Dowd. *Measuring Market Risk.* John Wiley & Sons, Ltd, 2005.

Fabrizio Durante, Wolfgang Härdle, Piotr Jaworski, and Tomasz Rychlik. *Copula Theory and Its Application.* Springer, 2009.

Colin M. Gallagher and Thomas J. Fisher. A comparison of various weighted portmanteau tests for time series goodness-of-fit. 2013.

Christian Genest, Ivan Kojadinovic, Johanna Nešlehova, and Jun Yan. A goodness-of-fit test for bivariate extreme-value copulas. *Bernoulli 17(1)*, 2011.

Alexios Ghalanos. Introduction to the rugarch package. 2014.

Ahmed Ghorbel and Abdelwahed Trabelsi. Measure of financial risk using conditional extreme value copulas with evt margins. *The Journal of Risk*, 2009.

Andreas Jakobsson. *An Introduction to Time Series Modeling.* Studentlitterartur, 2013.

Ribatet Mathieu and Sedki Mohammed. The t copula and related copulas. *International Statistical Review*, 2012.

Roger B. Nelson. *An Introduction to Copulas.* Springer, 2006.

Ruey S. Tsay. *Analysis of Financial Time Series*. John Wiley and Sons, Inc., Hoboken, New Jersey, 2005.

R. Vasudevay and J. Vasantha Kumari. On general error distributions. *Prob-Stat Forum*, 2013.

Tim Weithers. *Foreign Exchange - A Practical Guide to the FX Markets*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2006.

D. Wurtz, Y. Chalabi, and L. Luksan. Parameter estimation of arma models with garch/aparch errors an r an *Journal of Statistical Software*.

Jun Yan. Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 2007.